

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

### Search Topic:

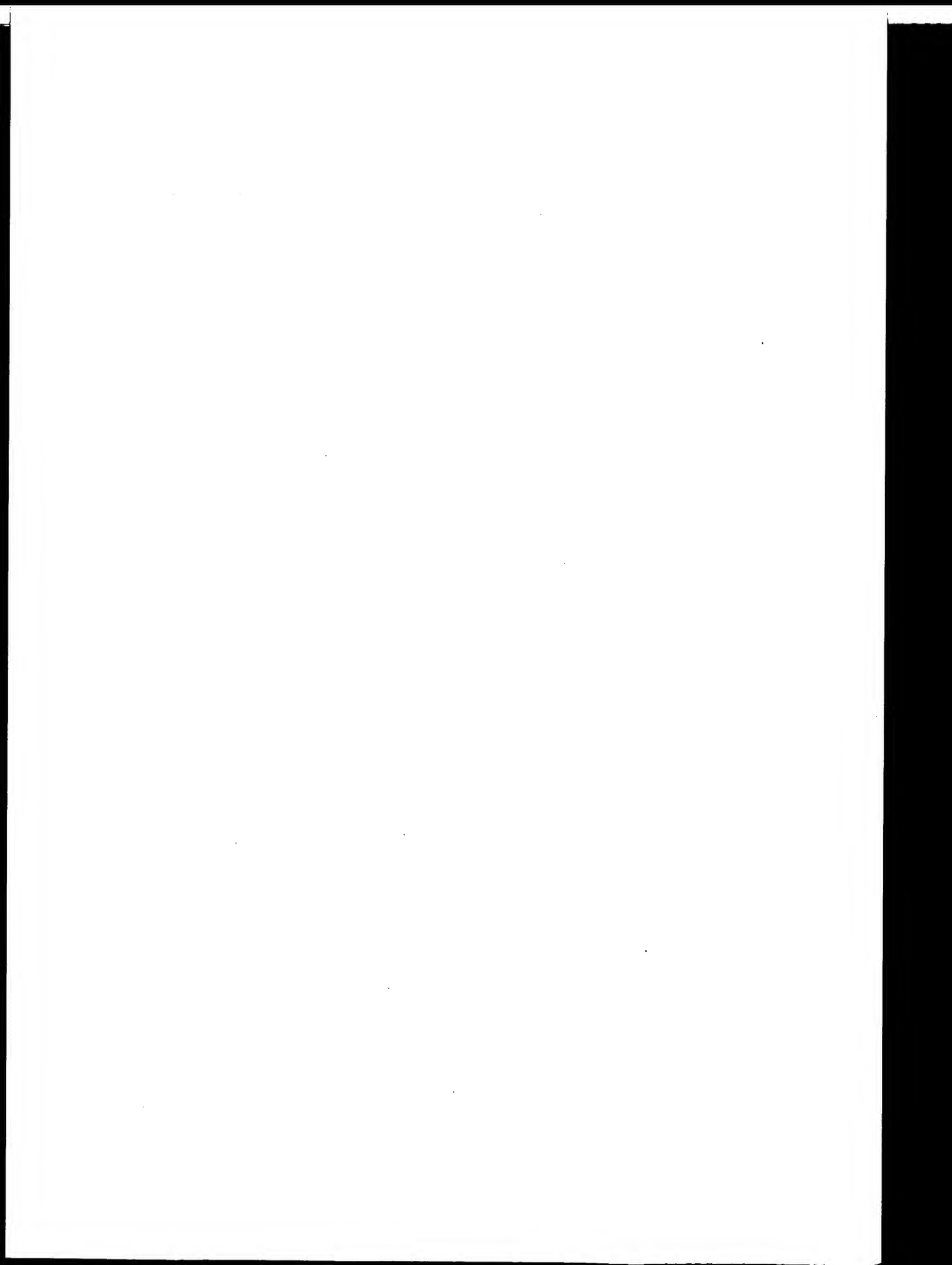
Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

### STAFF USE ONLY

Date completed: 03-25-03  
Searcher: Beverly E 4994  
Terminal time: \_\_\_\_\_  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: \_\_\_\_\_  
Number of Searches: \_\_\_\_\_  
Number of Databases: \_\_\_\_\_

Search Site  
\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S  
Type of Search  
\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

Vendors  
\_\_\_\_\_ IG Suite  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
✓ Other CGN





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:53:49 ; Search time 12426 Seconds  
(without alignments)  
11562.897 Million cell updates/sec

Title: US-10-092-880-3

Perfect score: 4937  
Sequence: 1 taaatacaagataataaa.....tttaacagggttattattatg 4937

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.bl.\*
- 9: gb.bl.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em\_sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4937	100.0	4937	6	AR032348	AR032348 Sequence
2	4937	100.0	4937	6	AR083955	AR083955 Sequence
3	4937	100.0	4937	6	AR145899	AR145899 Sequence
4	4937	100.0	4937	6	I25182	I25182 Sequence 3
5	4937	100.0	4937	6	I35771	I35771 Sequence 3
6	4935.4	100.0	4937	6	BD010356	BD010356 High molr
7	4900.2	99.3	8983	1	HIU08875	U08875 Haemophilus
8	4820.2	97.6	9323	6	AR032350	AR032350 Sequence
9	4820.2	97.6	9323	6	AR083957	AR083957 Sequence
10	4820.2	97.6	9323	6	AR145901	AR145901 Sequence
11	4820.2	97.6	9323	6	I35184	I25184 Sequence 6
12	4820.2	97.6	9323	6	I35773	I35773 Sequence 6
13	4817	97.6	9323	6	BD010358	BD010358 High molr
14	2689.8	54.5	9221	1	HIU08876	U08876 Haemophilus
15	2680.2	54.3	5116	6	AR032347	AR032347 Sequence
16	2680.2	54.3	5116	6	AR083954	AR083954 Sequence
17	2680.2	54.3	5116	6	AR145898	AR145898 Sequence
18	2680.2	54.3	5116	6	I25181	I25181 Sequence 1
19	2680.2	54.3	5116	6	I35770	I35770 Sequence 1
20	2567	52.0	9171	6	AR032349	AR032349 Sequence
21	2567	52.0	9171	6	AR083956	AR083956 Sequence
22	2567	52.0	9171	6	AR145900	AR145900 Sequence
23	2567	52.0	9171	6	I25183	I25183 Sequence 5
24	2559.8	51.8	5056	6	BD010355	BD010355 High molr
25	2507	50.8	9111	6	BD010357	BD010357 High molr
26	2368.6	48.0	6976	1	AF180944	AF180944 Haemophil
27	2246.6	45.5	4803	6	AR083959	AR083959 Sequence
28	2245	45.5	4803	6	BD010360	BD010360 High molr
29	2202.2	44.6	4702	6	AR032352	AR032352 Sequence
30	2202.2	44.6	4702	6	AR145903	AR145903 Sequence
31	2202.2	44.6	4702	6	I25186	I25186 Sequence 8
32	2202.2	44.6	4702	6	I35775	I35775 Sequence 8
33	1974	40.0	4794	6	AR083958	AR083958 Sequence
34	1959.8	39.7	4795	6	BD010359	BD010359 High molr
35	1476.4	29.9	4287	6	AR032351	AR032351 Sequence
36	1476.4	29.9	4287	6	AR145902	AR145902 Sequence
37	1476.4	29.9	4287	6	I25185	I25185 Sequence 7
38	1476.4	29.9	4287	6	I35774	I35774 Sequence 7
39	94.6	1.9	10029	1	AE013292	AE013292 Methanosa
C 40	83.8	1.7	177899	10	AC123042	AC123042 Mus muscu
C 41	83.4	1.7	325808	2	AC121787	AC121787 Mus muscu
C 42	80.8	1.6	210821	10	AL731814	AL731814 Mouse DNA
C 43	78.4	1.6	146734	10	AL590415	AL590415 Mouse DNA
C 44	78.2	1.6	190380	2	AC123847	AC123847 Mus muscu

## ALIGNMENTS

RESULT 1  
AR032348  
LOCUS AR032348 4937 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5869065.  
ACCESSION AR032348  
VERSION AR032348.1 GI:5947953  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 4937)  
AUTHORS Barenkamp,S.J. and St. Geme,J. William. III.  
TITLE High molecular weight surface proteins of non-typeable haemophilus  
JOURNAL Patent: US 5869065/A 3 09-FEB-1999;  
FEATURES Location/Qualifiers

Qy	961	ACTGTCGGTAAAGACGGCAGTGTAAATCTTTATTTGGTGGCAAGCTGAAACACGAGGTTGTG	102
Db	961	ACTGTCGGTAAAGACGGCAGTGTAAATCTTTATTTGGTGGCAAGCTGAAACACGAGGTTGTG	102
Qy	1021	ATTACGGTAAATTTGGTGGCAGCATTTCTTTTACTCGACGGCAAAAATACCATTCAGCGAT	1080
Db	1021	ATTACGGTAAATTTGGTGGCAGCATTTCTTTTACTCGACGGCAAAAATACCATTCAGCGAT	1080
Qy	1081	ATAATAAACCGAACCATTTACTTTTACAGCATTTGGCGCCCTGAAANTGAAGGGTCAATCTG	1140
Db	1081	ATAATAAACCCAACTTTACTTTTACAGCATTTGGCGCCCTGAAANTGAAGGGTCAATCTG	1140
Qy	1141	GCGGATATTTTGGCAAAAGCGCGGTAACTTAATGTCCGTGCTGCCACTATTTCGAAACCAA	1200
Db	1141	GCGGATATTTTGGCAAAAGCGCGGTAACTTAATGTCCGTGCTGCCACTATTTCGAAACCAA	1200
Qy	1201	GGTAACCTTTCGTGTGTCTGTAGCAAAAGATAAAAGCGGCAATATTGTCTTTCCGCG	1260
Db	1201	GGTAACCTTTCGTGTGTCTGTAGCAAAAGATAAAAGCGGCAATATTGTCTTTCCGCG	1260
Qy	1261	AAAGAGGTGAAGCGGGAATTTGGCGGTGTAAATTCGCGTCAAATCAGCAAGCTAAAGGC	1320
Db	1261	AAAGAGGTGAAGCGGGAATTTGGCGGTGTAAATTCGCGTCAAATCAGCAAGCTAAAGGC	1320
Qy	1321	GGCAAGCTGATGATTTACAGGGCGATAAAGTGCATTTAAAAACAGGTGCAGTTATTCGACCTT	1380
Db	1321	GGCAAGCTGATGATTTACAGGGCGATAAAGTGCATTTAAAAACAGGTGCAGTTATTCGACCTT	1380
Qy	1381	TCAGTTAAAGAAAGGGGAGAACTTTACCTTGGCGGTGACGAGCGGGCGGAAGGTAAAAAC	1440
Db	1381	TCAGTTAAAGAAAGGGGAGAACTTTACCTTGGCGGTGACGAGCGGGCGGAAGGTAAAAAC	1440
Qy	1441	GGCATTTCAATTTAGCAAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC	1500
Db	1441	GGCATTTCAATTTAGCAAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC	1500
Qy	1501	AAAGAAAAGCGGACGCGCTATTGTGTGGGCGATATTGCCGTTAATTGACGGCAATATT	1560
Db	1501	AAAGAAAAGCGGACGCGCTATTGTGTGGGCGATATTGCCGTTAATTGACGGCAATATT	1560
Qy	1561	AACGCTCAAGGTAGTGGTGATATCGCTAAAAACCGGTGTTTTGTGGAGACATCTGGGGCAT	1620
Db	1561	AACGCTCAAGGTAGTGGTGATATCGCTAAAAACCGGTGTTTTGTGGAGACATCTGGGGCAT	1620
Qy	1621	TATTTATCCATTGACAGCAATGCAATTTCTTAAACAAAAGAGTGGTCTAGACCCCTGAT	1680
Db	1621	TATTTATCCATTGACAGCAATGCAATTTCTTAAACAAAAGAGTGGTCTAGACCCCTGAT	1680
Qy	1681	GATGTAACAATTTGAAGCGGAAGACCCCTTCGCAATATACCGGTATAAATGATCAATTC	1740
Db	1681	GATGTAACAATTTGAAGCGGAAGACCCCTTCGCAATATACCGGTATAAATGATCAATTC	1740
Qy	1741	CCAACAGCCCGGTGAAGCAAGCGACCTTAANAANAATACGGACTCAAAACACGCTA	1800
Db	1741	CCAACAGCCCGGTGAAGCAAGCGACCTTAANAANAATACGGACTCAAAACACGCTA	1800
Qy	1801	ACCAATACACTATTTCCAAATTTCTGAAAACCGCTGGACAAATGATATACGGCATCA	1860
Db	1801	ACCAATACACTATTTCCAAATTTCTGAAAACCGCTGGACAAATGATATACGGCATCA	1860
Qy	1861	AGAAACTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT	1920
Db	1861	AGAAACTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT	1920
Qy	1921	AGTAAGGTCAGCTGGCGGAGCGGTTTCAGATTGATGGAGATATTACTTCTAAAGCGGA	1980
Db	1921	AGTAAGGTCAGCTGGCGGAGCGGTTTCAGATTGATGGAGATATTACTTCTAAAGCGGA	1980
Qy	1981	AATTTAACCAATTTATCTGGCGGATGGTTGATGTTTCATAAAAAATATTACGCTTGATCAG	2040
Db	1981	AATTTAACCAATTTATCTGGCGGATGGTTGATGTTTCATAAAAAATATTACGCTTGATCAG	2040
Y	2041	GGTTTTTTAAATATTACCGCGGCTTCGCTAGCTTTTGAAGGTGGAAATAACAAACACGCG	2100

Db 2041 GGTGTTTTTAAATATTACCGCGCTTCGGTAGCTTTTGAAGTGGAAATTAACAAAGCAGCG 2100  
QY 2101 GACGGGGCAATGCTAAATTTGTCGCCAGGCACTGTAAACCTTTACAGGAGAGGGA 2160  
Db 2101 GACGGGGCAATGCTAAATTTGTCGCCAGGCACTGTAAACCTTTACAGGAGAGGGA 2160  
QY 2161 GATTTACGGGCTAACACAGCTATCTTTAAACGGAACGGGTGAAGGTCTGAATATCATTTCA 2220  
Db 2161 GATTTACGGGCTAACACAGCTATCTTTAAACGGAACGGGTGAAGGTCTGAATATCATTTCA 2220  
QY 2221 TCAGTGAATTAATTTAACCCACCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280  
Db 2221 TCAGTGAATTAATTTAACCCACCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280  
QY 2281 ATTAACCAAACTACGAGAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGCACCTGG 2340  
Db 2281 ATTAACCAAACTACGAGAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGCACCTGG 2340  
QY 2341 AAGCTCAGTCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTATTAAATACATTTCA 2400  
Db 2341 AAGCTCAGTCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTATTAAATACATTTCA 2400  
QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGACGCTCTGCGGGGTGAATTTTAAAGCGC 2460  
Db 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGAGCTCTGAGGGGTGAATTTTAAAGCGC 2460  
QY 2461 GTAAATGGCAACATGTCTTAATCTCAAGAGAGGAGGAAAGTTAAATTTCAAAATTA 2520  
Db 2461 GTAAATGGCAACATGTCTTAATCTCAAGAGAGGAGGAAAGTTAAATTTCAAAATTA 2520  
QY 2521 CCAACGAGAACATGATCAACACAGCAACCTTTTACCAATTCGGTTTTTACCAATATCA 2580  
Db 2521 CCAACGAGAACATGATCAACACAGCAACCTTTTACCAATTCGGTTTTTACCAATATCA 2580  
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
Db 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
QY 2641 GAGTTAAATAGTGAATTAATATCTCAAGCGGCTAAATTTTACCTTTAAATTCCTCAT 2700  
Db 2641 GAGTTAAATAGTGAATTAATATCTCAAGCGGCTAAATTTTACCTTTAAATTCCTCAT 2700  
QY 2701 GTTCGGCGCATGACGCTTTTAAATCAACAAAGACTTAAACCATAAATGCAACCAATCA 2760  
Db 2701 GTTCGGCGCATGACGCTTTTAAATCAACAAAGACTTAAACCATAAATGCAACCAATCA 2760  
QY 2761 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGGTACGCGCAATGCGCATC 2820  
Db 2761 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGGTACGCGCAATGCGCATC 2820  
QY 2821 AATTTCAACCTACACATATCCATTTCTGGGGGTAATGTACACCTTGGTGGACAAAACCTCA 2880  
Db 2821 AATTTCAACCTACACATATCCATTTCTGGGGGTAATGTACACCTTGGTGGACAAAACCTCA 2880  
QY 2881 AGCAGCAGCATTACGGGGAATATTTACATTCGAGAAAGCAGCAATGTTAGCCTAGAAGCC 2940  
Db 2881 AGCAGCAGCATTACGGGGAATATTTACATTCGAGAAAGCAGCAATGTTAGCCTAGAAGCC 2940  
QY 2941 AATACGCCCCCTAATACGAAACATAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000  
Db 2941 AATACGCCCCCTAATACGAAACATAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000  
QY 3001 GTTAATGGAGTTTAACTTTAACTTGGCGAATTCAGATTTAAAGCAATCTCACTATT 3060  
Db 3001 GTTAATGGAGTTTAACTTTAACTTGGCGAATTCAGATTTAAAGCAATCTCACTATT 3060  
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAGACATAGATACCTTAATATACCGGCAATTTT 3120  
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAGACATAGATACCTTAATATACCGGCAATTTT 3120  
QY 3121 ACCAATAATGGCACTGGCGGAATTAATATAACAAAGAGGTGTTAAACCTTTGGCAATGTT 3180

Db 3121 ACCAATAATGGCACTGGCGGAATTAATATAACAAAGAGGTGTTAAACCTTTGGCAATGTT 3180  
QY 3181 ACCAATCATGTTGATTTTAAACATTTTACCACCTACAGCTAAACGCAACCAAGAGCATCATC 3240  
Db 3181 ACCAATCATGTTGATTTTAAACATTTTACCACCTACAGCTAAACGCAACCAAGAGCATCATC 3240  
QY 3241 GCGGAGATATAATCAACAAAAAGGAGCTTAAATATTACAGACAGTATAATATGATGCT 3300  
Db 3241 GCGGAGATATAATCAACAAAAAGGAGCTTAAATATTACAGACAGTATAATATGATGCT 3300  
QY 3301 GAAATCAAAATTTGGCGCAATATCTCGCAAAAAGAGCAACCTCAGATTTCTTCCGAT 3360  
Db 3301 GAAATCAAAATTTGGCGCAATATCTCGCAAAAAGAGCAACCTCAGATTTCTTCCGAT 3360  
QY 3361 AAAATTAATATCACCAACAGATTAACAAATCAAAAAGGGTATTGATGGAGAGGACTCTAGT 3420  
Db 3361 AAAATTAATATCACCAACAGATTAACAAATCAAAAAGGGTATTGATGGAGAGGACTCTAGT 3420  
QY 3421 TCAGATGCGACAAGTAATGCGCAACCTTAACCTATTAAAAACCAAGAAATTTGAAATTCACAGAA 3480  
Db 3421 TCAGATGCGACAAGTAATGCGCAACCTTAACCTATTAAAAACCAAGAAATTTGAAATTCACAGAA 3480  
QY 3481 GACCTAAGTATTTCAGGTTTCAATATAAGCAGAGATTACAGCCAAAGATGTTAGAGATTTA 3540  
Db 3481 GACCTAAGTATTTCAGGTTTCAATATAAGCAGAGATTACAGCCAAAGATGTTAGAGATTTA 3540  
QY 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTCCGGAAGCCAAACAGTAAGTAACTTTAAC 3600  
Db 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTCCGGAAGCCAAACAGTAAGTAACTTTAAC 3600  
QY 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGCTCACAATGACACTGAATAGCAAAAGTG 3660  
Db 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGCTCACAATGACACTGAATAGCAAAAGTG 3660  
QY 3661 AAACATCTAGCAGCAATGCGGCGAGTAAAGCAATACGCAACAGTACCGCTTAATC 3720  
Db 3661 AAACATCTAGCAGCAATGCGGCGAGTAAAGCAATACGCAACAGTACCGCTTAATC 3720  
QY 3721 AATTTGCAAAAAATTTAGAGTAACAAAGATATTTCTCTCTCAAAACAGTAATATC 3780  
Db 3721 AATTTGCAAAAAATTTAGAGTAACAAAGATATTTCTCTCTCAAAACAGTAATATC 3780  
QY 3781 ACCGGCTCGGAAAAAGGTTTACCACACAGCAGCTCGACCATTTAACGCAACAATGGCAAA 3840  
Db 3781 ACCGGCTCGGAAAAAGGTTTACCACACAGCAGCTCGACCATTTAACGCAACAATGGCAAA 3840  
QY 3841 GCAAGTATTACAAACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACAGGTAAGT 3900  
Db 3841 GCAAGTATTACAAACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACAGGTAAGT 3900  
QY 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATCCGCTCAAAATTTGAAGCAAAATCGGGT 3960  
Db 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATCCGCTCAAAATTTGAAGCAAAATCGGGT 3960  
QY 3961 GAGGCTTAATGTAACAGTCCAAACAGTCAATTTGGCGGTACAAATTTCCGGTAAATAGGTA 4020  
Db 3961 GAGGCTTAATGTAACAGTCCAAACAGTCAATTTGGCGGTACAAATTTCCGGTAAATAGGTA 4020  
QY 4021 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATGGCGCAAAATTAATGGGACA 4080  
Db 4021 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATGGCGCAAAATTAATGGGACA 4080  
QY 4081 GAAGGAGCTGCAACCTTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140  
Db 4081 GAAGGAGCTGCAACCTTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140  
QY 4141 ATCAGTTCAACTAAGGTGAGTACAGTCTTTGGCTCAGAAATGTTAGCTAGCATCGCAGGAAGC 4200  
Db 4141 ATCAGTTCAACTAAGGTGAGTACAGTCTTTGGCTCAGAAATGTTAGCTAGCATCGCAGGAAGC 4200  
QY 4201 ATTAATGCTGCTAATCTGACATTAATACTACAGGACCTTAAACACCGCTGGCAGGCTCG 4260  
Db 4201 ATTAATGCTGCTAATCTGACATTAATACTACAGGACCTTAAACACCGCTGGCAGGCTCG 4260



1081	QY	ATAATAAACCCAA	CCGATTTAC	GTACAGCATTTG	CGCGCGCTG	AAATGAA	CGGTC	CAATTCG	1140	
1081	Db	ATAATAAACCCAA	CCATTTAC	GTACAGCATTTG	CGCGCGCTG	AAATGAA	CGGTC	CAATTCG	1140	
1141	QY	GCGGATTTTTT	GCCAAAGCG	GTAACATTAAT	TGTCGGT	GCTGCC	ACTATT	TGAAACCAA	1200	
1141	Db	GCGGATTTTTT	GCCAAAGCG	GTAACATTAAT	TGTCGGT	GCTGCC	ACTATT	TGAAACCAA	1200	
1201	QY	GCTAAAC	TTCTGCTG	ATCTTGAAG	CAAAAGATA	AAAGCGG	CAATATG	TCTTTCCGCG	1260	
1201	Db	GCTAAAC	TTCTGCTG	ATCTTGAAG	CAAAAGATA	AAAGCGG	CAATATG	TCTTTCCGCG	1260	
1261	QY	AAAGAGG	TGAAGCG	GAAATTTG	CGCGTGTAT	TTCCGCT	CAAAATCAG	CAAGCTAAAGCG	1320	
1261	Db	AAAGAGG	TGAAGCG	GAAATTTG	CGCGTGTAT	TTCCGCT	CAAAATCAG	CAAGCTAAAGCG	1320	
1321	QY	GGCAAGCT	GATGATTAC	AGGCGATAA	AGTCACTTAAAA	ACAGGTG	CAGTTAT	TCGACCTT	1380	
1321	Db	GGCAAGCT	GATGATTAC	AGGCGATAA	AGTCACTTAAAA	ACAGGTG	CAGTTAT	TCGACCTT	1380	
1381	QY	TCAGGTAA	AGAGGGG	GAGAACTTAC	CTTGGCGGT	CACGAGCG	CGCGGAAG	GTAATAAAC	1440	
1381	Db	TCAGGTAA	AGAGGGG	GAGAACTTAC	CTTGGCGGT	CACGAGCG	CGCGGAAG	GTAATAAAC	1440	
1441	QY	GCGATTC	CAATTAG	CANAGAA	ACCTCTTTAG	AAAAAGG	CTCA	CCATCAATGATCAGCG	1500	
1441	Db	GCGATTC	CAATTAG	CANAGAA	ACCTCTTTAG	AAAAAGG	CTCA	CCATCAATGATCAGCG	1500	
1501	QY	AAAGCTCA	AGGTAG	TGTGATAT	CGCTAAAA	ACCGTGG	TTGTGG	GAGACATCGGGCAT	1560	
1501	Db	AAAGCTCA	AGGTAG	TGTGATAT	CGCTAAAA	ACCGTGG	TTGTGG	GAGACATCGGGCAT	1560	
1561	QY	AAAGCTCA	AGGTAG	TGTGATAT	CGCTAAAA	ACCGTGG	TTGTGG	GAGACATCGGGCAT	1620	
1561	Db	AAAGCTCA	AGGTAG	TGTGATAT	CGCTAAAA	ACCGTGG	TTGTGG	GAGACATCGGGCAT	1620	
1621	QY	TATTTAT	CCATTTG	CACAGCAAT	TGCTTAAAA	CAAAAGAG	TGGTGT	CTAGACCCCTGAT	1680	
1621	Db	TATTTAT	CCATTTG	CACAGCAAT	TGCTTAAAA	CAAAAGAG	TGGTGT	CTAGACCCCTGAT	1680	
1681	QY	GATGTAA	CAATTTG	AGCCGGA	AGACCCCTT	CGCAAT	AATAC	CCGSTATAAATGATGAATTC	1740	
1681	Db	GATGTAA	CAATTTG	AGCCGGA	AGACCCCTT	CGCAAT	AATAC	CCGSTATAAATGATGAATTC	1740	
1741	QY	CCAACG	CACCGGT	GAAAG	CAGGACCCCT	AAAAAAA	AATACG	GAACCTCAAAAA	CAACGCTA	1800
1741	Db	CCAACG	CACCGGT	GAAAG	CAGGACCCCT	AAAAAAA	AATACG	GAACCTCAAAAA	CAACGCTA	1800
1801	QY	ACCAATACA	CACTATT	TCAAAATTTAT	CTGAAAA	ACGCTTGG	ACAATGAAT	ATAACCGG	CATCA	1860
1801	Db	ACCAATACA	CACTATT	TCAAAATTTAT	CTGAAAA	ACGCTTGG	ACAATGAAT	ATAACCGG	CATCA	1860
1861	QY	AGAAAACT	TACGGTTAAT	TAGCTCAAT	CAACAT	TCGGAAG	CACCTCC	CACTTAATCTTCCAT	1920	
1861	Db	AGAAAACT	TACGGTTAAT	TAGCTCAAT	CAACAT	TCGGAAG	CACCTCC	CACTTAATCTTCCAT	1920	
1921	QY	AGTAAAG	TCAGCTG	CGGAGG	GTTCAGATTC	ATG	CGGATATT	ACTTCTTAAGCGG	1980	
1921	Db	AGTAAAG	TCAGCTG	CGGAGG	GTTCAGATTC	ATG	CGGATATT	ACTTCTTAAGCGG	1980	
1981	QY	AAATTTAA	CCANTTAT	CTCGGCG	ATGGTTC	ATGTCAT	AAAAATAT	TACGCTTGATCAG	2040	
1981	Db	AAATTTAA	CCANTTAT	CTCGGCG	ATGGTTC	ATGTCAT	AAAAATAT	TACGCTTGATCAG	2040	
2041	QY	GTTTTTT	TAATATTAC	CGCGCG	CTTCCGT	AGCTTTG	GAAGT	GAATAACA	AAAGCAGC	2100
2041	Db	GTTTTTT	TAATATTAC	CGCGCG	CTTCCGT	AGCTTTG	GAAGT	GAATAACA	AAAGCAGC	2100
2101	QY	GACGCG	CAATTCG	CTAAAAATTTG	CTGCC	ACAGG	CACTCT	TAACCAT	TACAGAGAGG	2160
2101	Db	GACGCG	CAATTCG	CTAAAAATTTG	CTGCC	ACAGG	CACTCT	TAACCAT	TACAGAGAGG	2160
2161	QY	GATTTTC	AGGGCTAA	CAACG	GTATCTTTAA	CCGGAAC	GGGTAA	GGTCTG	AAATATCATTTCA	2220

```
Db 3241 GCGGAGATATATCAACAAAAAGGAAGCTTAAATATTACAGACAGTAATATGATGCT 3300
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTTCACGATTTCTCCGAT 3360
Db 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTTCACGATTTCTCCGAT 3360
QY 3361 AAAATTAATATCACCACACAGATACAAATCAAAAGGGTATTGATGGAGGAGACTCTAGT 3420
Db 3361 AAAATTAATATCACCACACAGATACAAATCAAAAGGGTATTGATGGAGGAGACTCTAGT 3420
QY 3421 TCAGATGGCAGCAAGTAATGCCAACCTTAACATTTAAACCAAGAAATTTGAAATTTGACAGAA 3480
Db 3421 TCAGATGGCAGCAAGTAATGCCAACCTTAACATTTAAACCAAGAAATTTGAAATTTGACAGAA 3480
QY 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTATGAGATTTA 3540
Db 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTATGAGATTTA 3540
QY 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGTGCAGGCGCAAGCAACAGTAACTTTTAAAC 3600
Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGTGCAGGCGCAAGCAACAGTAACTTTTAAAC 3600
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAATGTGACACTAAATAAGCAAAAGTG 3660
Db 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAATGTGACACTAAATAAGCAAAAGTG 3660
QY 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAGCAATAGGCAACAGTACCGGCTTAACT 3720
Db 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAGCAATAGGCAACAGTACCGGCTTAACT 3720
QY 3721 ATTACTGCAAAAATGTAGAACTAAACAAAGATATTACTTCTCAAAACAGTAAATATC 3780
Db 3721 ATTACTGCAAAAATGTAGAACTAAACAAAGATATTACTTCTCAAAACAGTAAATATC 3780
QY 3781 ACCGCGTCGGAAGGTTACCAACACAGCGCTCGACCATTTAAGCAACAAATGGCAAA 3840
Db 3781 ACCGCGTCGGAAGGTTACCAACACAGCGCTCGACCATTTAAGCAACAAATGGCAAA 3840
QY 3841 GCAAGTATTACACCAACAGGCTGATATCAGCGGTAGCATTTCCGGTAAACACGGTAAGT 3900
Db 3841 GCAAGTATTACACCAACAGGCTGATATCAGCGGTAGCATTTCCGGTAAACACGGTAAGT 3900
QY 3901 GTTAGCGGACTGGTGTATTAACTAACTAAATCCGGCTCAAAATTTGAAGCGAAATCGGCT 3960
Db 3901 GTTAGCGGACTGGTGTATTAACTAACTAAATCCGGCTCAAAATTTGAAGCGAAATCGGCT 3960
QY 3961 GAGGCTAATGTAAACAGTGCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGGCTAATGTAAACAGTGCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020
QY 4021 AATGTTACGCAACAGCTGGCGATTACAGTTGGGAATGGCGGCAAGAAATTAATGCGACA 4080
Db 4021 AATGTTACGCAACAGCTGGCGATTACAGTTGGGAATGGCGGCAAGAAATTAATGCGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGCGTCAAGATGGTAGCATCGCAGGAAGC 4200
Db 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGCGTCAAGATGGTAGCATCGCAGGAAGC 4200
QY 4201 ATTAATGCTGCTAATGTGACATTAATTAATCAAGCAGCTTAACACCGTGGCAGGCTCG 4260
Db 4201 ATTAATGCTGCTAATGTGACATTAATTAATCAAGCAGCTTAACACCGTGGCAGGCTCG 4260
QY 4261 GATATTAAGCAACAGCGGCACTTGTTTATTAAACGCAAGAGATGCTTAAGCTAAATGGT 4320
Db 4261 GATATTAAGCAACAGCGGCACTTGTTTATTAAACGCAAGAGATGCTTAAGCTAAATGGT 4320
QY 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATGCATCAGGTGATAGTACAGAAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
```

---

```
QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATACATGCGGATTTAAACACAGTAAATGGTTA 4440
Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATACATGCGGATTTAAACACAGTAAATGGTTA 4440
QY 4441 AATATCATTTTCCAAAGATGGTGAACACACTGTGCGCTTAAGAGGCAAGAAATTTGAGTG 4500
Db 4441 AATATCATTTTCCAAAGATGGTGAACACACTGTGCGCTTAAGAGGCAAGAAATTTGAGTG 4500
QY 4501 AAATATATCAGCCAGGCTGACAGTGTAGAGAAGTAAATGAACGGAACCGCTCCTT 4560
Db 4501 AAATATATCAGCCAGGCTGACAGTGTAGAGAAGTAAATGAACGGAACCGCTCCTT 4560
QY 4561 GAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACAACTTAGCTAAACTTGGGTAAAGT 4620
Db 4561 GAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACAACTTAGCTAAACTTGGGTAAAGT 4620
QY 4621 GCTGTACGTTTTTGTGAGCCAAATTAATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
Db 4621 GCTGTACGTTTTTGTGAGCCAAATTAATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
QY 4681 ACCAGACCGCTCAAGTCAAGTGATATTTCTGAAGGTAAGCGGTGTTTCTCAAGTGGTAAT 4740
Db 4681 ACCAGACCGCTCAAGTCAAGTGATATTTCTGAAGGTAAGCGGTGTTTCTCAAGTGGTAAT 4740
QY 4741 GCGCAGCAGTATGTACCAATGTTGCTGACGATGGACACCGCTAGTCAATTTGACAAG 4800
Db 4741 GCGCAGCAGTATGTACCAATGTTGCTGACGATGGACACCGCTAGTCAATTTGACAAG 4800
QY 4801 GTAGATTTCATCCTGCAATGAAGTCAATTTATTTTCGTTATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTCATCCTGCAATGAAGTCAATTTATTTTCGTTATTTACTGTGGGTTAAA 4860
QY 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAAATAAAGTATTTT 4920
Db 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAAATAAAGTATTTT 4920
QY 4921 AACAGGTTATTATTATG 4937
Db 4921 AACAGGTTATTATTATG 4937

RESULT 3
ARI45899
LOCUS ARI45899
DEFINITION Sequence 3 from patent US 6218141.
ACCESSION ARI45899
VERSION ARI45899.1 GI:15109088
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4937)
AUTHORS Barenkamp,S.J.
TITLE High molecular weight surface proteins of non-typeable haemophilus
JOURNAL Patent: US 6218141-A 3 17-APR-2001;
FEATURES Location/Qualifiers
source 1..4937
BASE COUNT 1729 a 948 c 1009 g 1251 t
ORIGIN

Query Match 100.0%; Score 4937; DB 6; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATATACAGATTAATAAATAAATCAAGATTTTGTGTGACAAACAAACAATTTACAA 60
Db 1 TAAATATACAGATTAATAAATAAATCAAGATTTTGTGTGACAAACAAACAATTTACAA 60
QY 61 CACCTTTTTTTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
Db 61 CACCTTTTTTTCAGTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
```





[illegible]

Db	3361	AAATTTAAATATCACCAAAACAGATAACAATCAAAAAGGGTATTGATGAGAGGACTCTAGT	3420
QY	3421	TCAGATTGCGACAAGTAATGCGCAACCTAACTATTAAACCAAGAAGTTGAATTTGACAGAA	3480
Db	3421	TCAGATTGCGACAAGTAATGCGCAACCTAACTATTAAACCAAGAAGTTGAATTTGACAGAA	3480
QY	3481	GACCTAAGTATTTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA	3540
Db	3481	GACCTAAGTATTTTCAGGTTTCAATAAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTA	3540
QY	3541	ACTATTGGCAACAGTAAATGACGGTAAACAGGGTGCCGAAGCCCAAAACAGTAACCTTTTAAC	3600
Db	3541	ACTATTGGCAACAGTAAATGACGGTAAACAGGGTGCCGAAGCCCAAAACAGTAACCTTTTAAC	3600
QY	3601	AATGTTAAAGATTCAAAAATCTCTGCTGACGGTGACAATGTGACACTAAATACCAAGTG	3660
Db	3601	AATGTTAAAGATTCAAAAATCTCTGCTGACGGTGACAATGTGACACTAAATACCAAGTG	3660
QY	3661	AAAACATCTAGCAGCAATGGCGACGTGAAGCAATAGCGACAACGATACCGGCTTAACT	3720
Db	3661	AAAACATCTAGCAGCAATGGCGACGTGAAGCAATAGCGACAACGATACCGGCTTAACT	3720
QY	3721	ATTACTGCAAAAATAATGTAAAGTTAAACAAAGATTATTACTTCTCTCAAAACAGTAAATATC	3780
Db	3721	ATTACTGCAAAAATAATGTAAAGTTAAACAAAGATTATTACTTCTCTCAAAACAGTAAATATC	3780
QY	3781	ACCGGTCGGAAAGGTTACACACACAGCAGGCTGCACCATTAACGCAACAAATGCAAA	3840
Db	3781	ACCGGTCGGAAAGGTTACACACACAGCAGGCTGCACCATTAACGCAACAAATGCAAA	3840
QY	3841	GCAAGTATTACAACAAAACAGAGTGATATCAGCGGTACGATTTCGCGTAAACAGGTAAGT	3900
Db	3841	GCAAGTATTACAACAAAACAGAGTGATATCAGCGGTACGATTTCGCGTAAACAGGTAAGT	3900
QY	3901	GTTACGGGACTGGTGATTTTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGT	3960
Db	3901	GTTACGGGACTGGTGATTTTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGT	3960
QY	3961	GAGGCTAATGTAAACAAGTGCAACAGTGACAAATTTGGCGGTACAATTTCCGGTAAATACGTTA	4020
Db	3961	GAGGCTAATGTAAACAAGTGCAACAGTGACAAATTTGGCGGTACAATTTCCGGTAAATACGTTA	4020
QY	4021	AATGTTAGCGAAACGCTGGCGATTTTAACAGTGGGAATGGCGCAGAAATTAATCGACA	4080
Db	4021	AATGTTAGCGAAACGCTGGCGATTTTAACAGTGGGAATGGCGCAGAAATTAATCGACA	4080
QY	4081	GAAGGAGCTGCAACCTTAACGCAACAGGGAATCCTTGACTACTCAAGCCGTTCTAGC	4140
Db	4081	GAAGGAGCTGCAACCTTAACGCAACAGGGAATCCTTGACTACTCAAGCCGTTCTAGC	4140
QY	4141	ATCACTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATCGAGCAAGC	4200
Db	4141	ATCACTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATCGAGCAAGC	4200
QY	4201	ATTAATGCTGCTAATGTACATTAATACTACAGGCACCTTAACCAACCGTGGCAGGCTCG	4260
Db	4201	ATTAATGCTGCTAATGTACATTAATACTACAGGCACCTTAACCAACCGTGGCAGGCTCG	4260
QY	4261	GATATTAAGCAACCAGCGCACCTTGGTTATTAAACGCAAAAGATGCTAAAGTAAATGGT	4320
Db	4261	GATATTAAGCAACCAGCGCACCTTGGTTATTAAACGCAAAAGATGCTAAAGTAAATGGT	4320
QY	4321	GATGCATCAGGTGATGTACAGAGTGAATGCAGTCAACGCAAGGGCTCTGGTAGTGTG	4380
Db	4321	GATGCATCAGGTGATGTACAGAGTGAATGCAGTCAACGCAAGGGCTCTGGTAGTGTG	4380
QY	4381	ACTCGGCAACCTCAAGCAGTGCTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA	4440
Db	4381	ACTCGGCAACCTCAAGCAGTGCTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA	4440
QY	4441	AATATCATTTGCAAGATGGTAGAACAACACTGTGGCTTTAAGAGGCAAGAAATTCAGGTG	4500
Db	4441	AATATCATTTGCAAGATGGTAGAACAACACTGTGGCTTTAAGAGGCAAGAAATTCAGGTG	4500





QY 1321 GCGAAGCTGATGATTACAGCGGATAAGTACATTAATAAAGAGTGCAGTTATCGACCTT 1380  
DB 1321 GCGAAGCTGATGATTACAGCGGATAAGTACATTAATAAAGAGTGCAGTTATCGACCTT 1380  
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACAGCGCGGAGGTAATAAAC 1440  
DB 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACAGCGCGGAGGTAATAAAC 1440  
QY 1441 GGCATTCAATTAGCAAGAAAGCTCTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500  
DB 1441 GGCATTCAATTAGCAAGAAAGCTCTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500  
QY 1501 AAGAAAAAGCGGAGCGCGTATTTGTGGGCGATATTTGCGTTAAATTGACGCGCAATATT 1560  
DB 1501 AAGAAAAAGCGGAGCGCGTATTTGTGGGCGATATTTGCGTTAAATTGACGCGCAATATT 1560  
QY 1561 AACGCTCAAGGTAGTGGTATGATGCGTAAACCGGTGGTTTGTGGAGACATCGGGGAT 1620  
DB 1561 AACGCTCAAGGTAGTGGTATGCGTAAACCGGTGGTTTGTGGAGACATCGGGGAT 1620  
QY 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAAAACAAAAAGAGTGGTTGTAGACCCCTGAT 1680  
DB 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAAAACAAAAAGAGTGGTTGTAGACCCCTGAT 1680  
QY 1681 GATGTAAACAAATTGAAGCCGAAGACCCCTTCGCAATATATACCGGTATAAATGATGAATTC 1740  
DB 1681 GATGTAAACAAATTGAAGCCGAAGACCCCTTCGCAATATATACCGGTATAAATGATGAATTC 1740  
QY 1741 CCACAGGCGCGGTGAAGCAAGCGACCCCTTAAAAAATAGGAATCAAAACACGCTA 1800  
DB 1741 CCACAGGCGCGGTGAAGCAAGCGACCCCTTAAAAAATAGGAATCAAAACACGCTA 1800  
QY 1801 ACCAATACAACTATTTCAAATTTATCTGAAAAAGCGCTGGCAATGAATATACGCGCATCA 1860  
DB 1801 ACCAATACAACTATTTCAAATTTATCTGAAAAAGCGCTGGCAATGAATATACGCGCATCA 1860  
QY 1861 AGAAAACTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920  
DB 1861 AGAAAACTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920  
QY 1921 AGTAAAGTACAGCGTGGGAGGCGCTTGCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980  
DB 1921 AGTAAAGTACAGCGTGGGAGGCGCTTGCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980  
QY 1981 AATTTAAACATTTATCTGCGGAGTGGTGTGATTTGATGAGATATTACTTCTAAAGCGGA 2040  
DB 1981 AATTTAAACATTTATCTGCGGAGTGGTGTGATTTGATGAGATATTACTTCTAAAGCGGA 2040  
QY 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATATAACAAAGCACGC 2100  
DB 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATATAACAAAGCACGC 2100  
QY 2101 GACGGGCAATGCTAAATTTGTCGCCAGGCGACTGTAAACATTACAGGAGGGGAAAA 2160  
DB 2101 GACGGGCAATGCTAAATTTGTCGCCAGGCGACTGTAAACATTACAGGAGGGGAAAA 2160  
QY 2161 GATTTACGGGCTAACAGGTATCTTTAAACGGAAGCGGTAAAGTCTGAATATCATTTCA 2220  
DB 2161 GATTTACGGGCTAACAGGTATCTTTAAACGGAAGCGGTAAAGTCTGAATATCATTTCA 2220  
QY 2221 TCAGTGAATAATTAAACCCACAATCTTAGTGGCAACAATTAAACATATCTGGGAATATAACA 2280  
DB 2221 TCAGTGAATAATTAAACCCACAATCTTAGTGGCAACAATTAAACATATCTGGGAATATAACA 2280  
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTTGGCAAAACAGCCATGATTCGCACTGG 2340  
DB 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTTGGCAAAACAGCCATGATTCGCACTGG 2340  
QY 2341 AACGTCAGTGTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTATTAATACATTTCA 2400  
DB 2341 AACGTCAGTGTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTATTAATACATTTCA 2400  
QY 2401 AGCAATAGCAAGGCTTAACACACAGATATAGAAGCTCTCGAGGGGTGAATTTTAAACGGC 2460

DB 2401 AGCAATAGCAAGGCTTAACACACAGATATAGAAGCTCGAGGGTGAATTTTAAACGGC 2460  
QY 2461 GTAAATGGCAACATGCTCAATCAATCTCAAGAGAGGAGGAAAGTTAAATTTCAAAATAAAA 2520  
DB 2461 GTAAATGGCAACATGCTCAATCAATCTCAAGAGAGGAGGAAAGTTAAATTTCAAAATAAAA 2520  
QY 2521 CCAACAGAGAACATGAACACAGCAAAACCTTTACCAATTCGGTTTTAGCCAAATATCACA 2580  
DB 2521 CCAACAGAGAACATGAACACAGCAAAACCTTTACCAATTCGGTTTTAGCCAAATATCACA 2580  
QY 2581 GCCACTGTGGGGCTCTGTTTTTTTTCATATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
DB 2581 GCCACTGTGGGGCTCTGTTTTTTTTCATATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACCGCGTAAATTTTACCTTAAATTTCCCAT 2700  
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACCGCGTAAATTTTACCTTAAATTTCCCAT 2700  
QY 2701 GTTCGCGCGATGAGCTTTTAAAAATCAACAAAGACTTAACCATTAATGCAACCAATTC 2760  
DB 2701 GTTCGCGCGATGAGCTTTTAAAAATCAACAAAGACTTAACCATTAATGCAACCAATTC 2760  
QY 2761 AATTTCAAGCTTCAGACAGAGCAAGAGATGTTTTTATGACGGGTACGCAACCAATGCGCATC 2820  
DB 2761 AATTTCAAGCTTCAGACAGAGCAAGAGATGTTTTTATGACGGGTACGCAACCAATGCGCATC 2820  
QY 2821 AATTTCAAGCTTCAGACAGAGCAAGAGATGTTTTTATGACGGGTAAATGTCACCCCTTGGTGGACAAACTCA 2880  
DB 2821 AATTTCAAGCTTCAGACAGAGCAAGAGATGTTTTTATGACGGGTAAATGTCACCCCTTGGTGGACAAACTCA 2880  
QY 2881 AGCAGCAGCATTCAGGGGAATATTACTATCGAGAAAGCAACCAATGTTAGCTTAGAAGCC 2940  
DB 2881 AGCAGCAGCATTCAGGGGAATATTACTATCGAGAAAGCAACCAATGTTAGCTTAGAAGCC 2940  
QY 2941 AATAACCCCTTAATCAGCAAAACATAAAGGATAGAGTTATAAACTTGGCAGCTTGCTC 3000  
DB 2941 AATAACCCCTTAATCAGCAAAACATAAAGGATAGAGTTATAAACTTGGCAGCTTGCTC 3000  
QY 3001 GTTAATGGGAGTTTAACTTAACTGGCAAAATGCAGATATTAAAGCAATCTCAGTATT 3060  
DB 3001 GTTAATGGGAGTTTAACTTAACTGGCAAAATGCAGATATTAAAGCAATCTCAGTATT 3060  
QY 3061 TCAGAAAGCCCTTAACTTAAAGGAAAGACTAGAGATACCCCTAAATATCAGCGCAATTTT 3120  
DB 3061 TCAGAAAGCCCTTAACTTAAAGGAAAGACTAGAGATACCCCTAAATATCAGCGCAATTTT 3120  
QY 3121 ACCAATAATGGGCTGCGGAAATTAATAACAAGGAGTGGTAAACCTTGGCAATGTT 3180  
DB 3121 ACCAATAATGGGCTGCGGAAATTAATAACAAGGAGTGGTAAACCTTGGCAATGTT 3180  
QY 3181 ACCAATGATGGTCAATTAACATTAACCTCAGCTTAACCGCAACCAAGAGCAATCATC 3240  
DB 3181 ACCAATGATGGTCAATTAACATTAACCTCAGCTTAACCGCAACCAAGAGCAATCATC 3240  
QY 3241 GCGGAGATATAATCAACAAAGGAAAGCTTAAATATTACAGACAGTAAATGATGCT 3300  
DB 3241 GCGGAGATATAATCAACAAAGGAAAGCTTAAATATTACAGACAGTAAATGATGCT 3300  
QY 3301 GAATTCCAATTTGGCGGCAATATCTCGGCAAAAGAGGCAACCTCAGGATTTCTTCCGAT 3360  
DB 3301 GAATTCCAATTTGGCGGCAATATCTCGGCAAAAGAGGCAACCTCAGGATTTCTTCCGAT 3360  
QY 3361 AAAATTAATATCACCAACAGATATACTTCAAAAGGATTTGATGGAGAGGACTCTAGT 3420  
DB 3361 AAAATTAATATCACCAACAGATATACTTCAAAAGGATTTGATGGAGAGGACTCTAGT 3420  
QY 3421 TCAGATGCGCAAGTAAATGCCAACCTTAATTTAAACCAAAAGAAATGAAATTTGACAAA 3480  
DB 3421 TCAGATGCGCAAGTAAATGCCAACCTTAATTTAAACCAAAAGAAATGAAATTTGACAAA 3480  
QY 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACGCCAAAGATGTTAGAGATTTA 3540





|||||  
Db 2521 CCAACGAGACNATGAACACAGCAACCTTTACCAATTCGTTTTAGCCAAATATCACA 2580  
QY 2581 GCACATGGTGGGGCTCTCTTTTTTTTATATATATATGCAACCAATCTGGCAGAGGGCT 2640  
Db 2581 GCACATGGTGGGGCTCTCTTTTTTTTATATATATATGCAACCAATCTGGCAGAGGGCT 2640  
QY 2641 GAGTTAAAATAGTGAATTAATATCTCTACGGCGCTTAATTTTACCTTAATTTCCCAT 2700  
Db 2641 GAGTTAAAATAGTGAATTAATATCTCTACGGCGCTTAATTTTACCTTAATTTCCCAT 2700  
QY 2701 GTTCGGGGATGACGCTTTTAAAATCAACAAAGACTTAACCAATAAATGCAACCAATCA 2760  
Db 2701 GTTCGGGGATGACGCTTTTAAAATCAACAAAGACTTAACCAATAAATGCAACCAATCA 2760  
QY 2761 AATTTACGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCCATC 2820  
Db 2761 AATTTACGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCCATC 2820  
QY 2821 AATTTCAACCTTACACATATCCATCTCTGGCGGTAATGTCACCCCTTGGTGGCAAAAATCA 2880  
Db 2821 AATTTCAACCTTACACATATCCATCTCTGGCGGTAATGTCACCCCTTGGTGGCAAAAATCA 2880  
QY 2881 AGCAGCAGCATTACGGGGATATATCTATCGAGAAAGCAGCAATCTTACGCTAGAGCC 2940  
Db 2881 AGCAGCAGCATTACGGGGATATATCTATCGAGAAAGCAGCAATCTTACGCTAGAGCC 2940  
QY 2941 AATAAGCCCTTAATCAGCAAAACATATAAGGGATAGAGTTATATAAACTTGGCAGCTTGCTC 3000  
Db 2941 AATAAGCCCTTAATCAGCAAAACATATAAGGGATAGAGTTATATAAACTTGGCAGCTTGCTC 3000  
QY 3001 GTTAATGGAGTTTAAAGTTTAACTTGGCGAAATGCGAGATATATAAGGCAATCTCACTAT 3060  
Db 3001 GTTAATGGAGTTTAAAGTTTAACTTGGCGAAATGCGAGATATATAAGGCAATCTCACTAT 3060  
QY 3061 TCAGAAAGCGCCACTTTTAAAGAAAGACTTAGAGATACCCCTTAATATACCGCGCAATTTT 3120  
Db 3061 TCAGAAAGCGCCACTTTTAAAGAAAGACTTAGAGATACCCCTTAATATACCGCGCAATTTT 3120  
QY 3121 ACCAATATGCACTGCGGAAATTAATATACCAAGAGTGGTAAACCTTTGGCAATGTT 3180  
Db 3121 ACCAATATGCACTGCGGAAATTAATATACCAAGAGTGGTAAACCTTTGGCAATGTT 3180  
QY 3181 ACCAATGATGATTTAAACATTAACCATTAACCAAGAGTGGTAAACCTTTGGCAATGTT 3240  
Db 3181 ACCAATGATGATTTAAACATTAACCATTAACCAAGAGTGGTAAACCTTTGGCAATGTT 3240  
QY 3241 GCGGAGATATATCAACAAAGAGGAGCTTAATATTAACAGACATTAATATGATGCT 3300  
Db 3241 GCGGAGATATATCAACAAAGAGGAGCTTAATATTAACAGACATTAATATGATGCT 3300  
QY 3301 GAAATCCAAATGCGGCAATATCTCGCAAAAGAGGCAACCTTACGATTTCTCCGAT 3360  
Db 3301 GAAATCCAAATGCGGCAATATCTCGCAAAAGAGGCAACCTTACGATTTCTCCGAT 3360  
QY 3361 AAAATTAATATCAACCAAGATTAACCAATCAAAAGGGTATTTGATGAGAGGACTCTAGT 3420  
Db 3361 AAAATTAATATCAACCAAGATTAACCAATCAAAAGGGTATTTGATGAGAGGACTCTAGT 3420  
QY 3421 TCAGATGCAAGAAAGTAAATGCAACCTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480  
Db 3421 TCAGATGCAAGAAAGTAAATGCAACCTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480  
QY 3481 GACCTAAGTATTTACAGTTTCAATTAAGACAGATTAACAGCAAGATGTTAGAGATTTA 3540  
Db 3481 GACCTAAGTATTTACAGTTTCAATTAAGACAGATTAACAGCAAGATGTTAGAGATTTA 3540  
QY 3541 ACTATTGCAACAGTAAATGAGCGGTAAACAGCGGTGCGGAGCCCAACAGTAACCTTTAAC 3600  
Db 3541 ACTATTGCAACAGTAAATGAGCGGTAAACAGCGGTGCGGAGCCCAACAGTAACCTTTAAC 3600  
QY 3601 AATGTTAAAGATTTCAAAATCTCTGCTGACCGGTACCAATGTGACACTAAATAGCAAGTG 3660  
|||||

Db 3601 AATGTTAAAGATTTCAAAATCTCTGCTGACGGTCACAAATGTGACACTAAATAGCAAGTG 3660  
QY 3661 AAAACATCTAGCAGCAATGCGGACGTGAAAGCAATAGCGCAACAGATACCGGCTTAAC 3720  
Db 3661 AAAACATCTAGCAGCAATGCGGACGTGAAAGCAATAGCGCAACAGATACCGGCTTAAC 3720  
QY 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATC 3780  
Db 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCTCAAAACAGTAAATATC 3780  
QY 3781 ACCGGTTCGGAAGAGTTTACCACACAGCAGCTCGACATTAACGCAACAAATGGCAAA 3840  
Db 3781 ACCGGTTCGGAAGAGTTTACCACACAGCAGCTCGACATTAACGCAACAAATGGCAAA 3840  
QY 3841 GCAAGTATTACACCAAAACAGGTGATATCAGCGGTAGATTTCCGGTAACACGTAAGT 3900  
Db 3841 GCAAGTATTACACCAAAACAGGTGATATCAGCGGTAGATTTCCGGTAACACGTAAGT 3900  
QY 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATTCGGGT 3960  
Db 3901 GTTAGCGGACTGGTGAATTTAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATTCGGGT 3960  
QY 3961 GAGGCTAATGTAAAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAATACGTA 4020  
Db 3961 GAGGCTAATGTAAAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAATACGTA 4020  
QY 4021 AATGTTACGGCAACGCTGGCGATTTAAACAGTTGGGAATGGCGCAAAATTAATCGACA 4080  
Db 4021 AATGTTACGGCAACGCTGGCGATTTAAACAGTTGGGAATGGCGCAAAATTAATCGACA 4080  
QY 4081 GAAGGAGCTGCAACCTTTAAACCGCAACAGGAATACCTTGCAGTACGAGCGGTTCTAGC 4140  
Db 4081 GAAGGAGCTGCAACCTTTAAACCGCAACAGGAATACCTTGCAGTACGAGCGGTTCTAGC 4140  
QY 4141 ATCAGCTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGATGTTAGATGTCAGCAAGC 4200  
Db 4141 ATCAGCTTCAACTAAGGGTCAGGTAGACCTCTTGGCTCAGATGTTAGATGTCAGCAAGC 4200  
QY 4201 ATTAATGCTGCTAATGTGACATTAATACAGGCACTTAAACCAAGCTTAAACCAAGCTG 4260  
Db 4201 ATTAATGCTGCTAATGTGACATTAATACAGGCACTTAAACCAAGCTTAAACCAAGCTG 4260  
QY 4261 GATATTAAGCAACACAGCGGCACCTTGGTTATTAACGCAAAAGATGCTAAGCTAATGTT 4320  
Db 4261 GATATTAAGCAACACAGCGGCACCTTGGTTATTAACGCAAAAGATGCTAAGCTAATGTT 4320  
QY 4321 GATGCAATCAGGTGATGATGACAGAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4380  
Db 4321 GATGCAATCAGGTGATGATGACAGAGTGAATGCAAGTGAATGCAAGTGAATGCAAGTGA 4380  
QY 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCCTGGGATTTAAACACAGTAAATGGGTTA 4440  
Db 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCCTGGGATTTAAACACAGTAAATGGGTTA 4440  
QY 4441 AATATCATTTTGAAGATGTTAGAACACACTGCGGTTAAGAGCAAGCAATTTAGAGTG 4500  
Db 4441 AATATCATTTTGAAGATGTTAGAACACACTGCGGTTAAGAGCAAGCAATTTAGAGTG 4500  
QY 4501 AATATATATCAGCCAGGTGTAGCAAGTGTAGAGAAAGTAAATTTAGCGGAAACCGTCTCT 4560  
Db 4501 AATATATATCAGCCAGGTGTAGCAAGTGTAGAGAAAGTAAATTTAGCGGAAACCGTCTCT 4560  
QY 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAAACTTGGTGTAAAGT 4620  
Db 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAAACTTGGTGTAAAGT 4620  
QY 4621 GCTGTACGTTTTTGTGAGCCAAATTAATACAAATTTACAGTCAATACACAAATTAATTTACA 4680  
Db 4621 GCTGTACGTTTTTGTGAGCCAAATTAATACAAATTTACAGTCAATACACAAATTAATTTACA 4680  
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAAGTAAAGCGGTTCTTCAAGTGTAAAT 4740  
Db 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAAGTAAAGCGGTTCTTCAAGTGTAAAT 4740



```

QY 4741 GCGCAGGATGTACCAATGTGCTGACGATGGACGCCGTAGTCACTAATTGACAAG 4800
|
|
|
Db 4741 GCGCAGGATGTACCAATGTGCTGACGATGGACGCCGTAGTCACTAATTGACAAG 4800
|
|
|
QY 4801 GTAGATTTTCATCTGCAATGAAGTCACTTTTATTTTCGTAATTTACTGTGTGGGTTAAA 4860
|
|
|
Db 4801 GTAGATTTTCATCTGCAATGAAGTCACTTTTATTTTCGTAATTTACTGTGTGGGTTAAA 4860
|
|
|
QY 4861 GTTCTAGTGGGCTTTACCATCTTTGTAATAAATACGGAGAAATACAATAAAGTATTTT 4920
|
|
|
Db 4861 GTTCTAGTGGGCTTTACCATCTTTGTAATAAATACGGAGAAATACAATAAAGTATTTT 4920
|
|
|
QY 4921 AACAGTTTATTATTATG 4937
|
|
|
Db 4921 AACAGTTTATTATTATG 4937
|
|
|
RESULT 6
BD010356
LOCUS
DEFINITION High molecular weight surface protein of non-typeable haemophils.
ACCESSION BD010356
VERSION BD010356.1 GI:18638729
KEYWORDS JP 2001503602-A/2.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 4937)
AUTHORS Barenkamp,S.J.
TITLE High molecular weight surface protein of non-typeable haemophils
JOURNAL Patent: JP 2001503602-A 21-MAR-2001;
SENTLOUIS UNIV, WASHINGTON UNIV
COMMENT OS Haemophils
PN JP 2001503602-A/2
PD 21-MAR-2001
PR 01-APR-1997 JP 1997535346
PR 01-APR-1996 US 08/617697
PI STEPHEN J BARENKAMP
PC C07H21/02,C07H21/04,C12P21/06,A61K39/102
CC
FH
FT
FT
FEATURES
source Location/Qualifiers
1..4937
/organism="Haemophils".
Location/Qualifiers
1..4937
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1728 a 949 c 1009 g 1251 t
ORIGIN

Query Match 100.0%; Score 4935.4; DB 6; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4936; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAAATATACAGATATAAATAAATCAAGATTTTGTGATGACAAACCAACAAATTACAA 60
|
|
|
Db 1 TAAATATACAGATATAAATAAATCAAGATTTTGTGATGACAAACCAACAAATTACAA 60
|
|
|
QY 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATAAAACCGCCATATAA 120
|
|
|
Db 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATAAAACCGCCATATAA 120
|
|
|
QY 121 AATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
|
|
|
Db 121 AATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180
|
|
|
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
|
|
|
Db 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
|
|
|
QY 241 CACATGAATGATCAACCGAGGAAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
|
|
|

```

```

Db 241 CACATGAATGATCAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
|
|
|
QY 301 CAAAGCAATGATTAAGTAATTTTAAATTTTCAACTAACCCTTAGGAGAAAATATGAAACAAG 360
|
|
|
Db 301 CAAAGCAATGATTAAGTAATTTTAAATTTTCAACTAACCCTTAGGAGAAAATATGAAACAAG 360
|
|
|
QY 361 ATATATGCTGTCAAAATTCAGCAAAACGCTGAATGCTTTGGTTGCTGTCTGTAATTTGGCA 420
|
|
|
Db 361 ATATATGCTGTCAAAATTCAGCAAAACGCTGAATGCTTTGGTTGCTGTCTGTAATTTGGCA 420
|
|
|
QY 421 CGGGTGTGTGACCAATCCACAGAAAAGGCTTCGGCTATGTTACTATCTTTTAGGTGTAAAC 480
|
|
|
Db 421 CGGGTGTGTGACCAATCCACAGAAAAGGCTTCGGCTATGTTACTATCTTTTAGGTGTAAAC 480
|
|
|
QY 481 CACTTAGCGTTAAAGCCACTTTTCGGCTATGTTACTATCTTTTAGGTGTAAAC 540
|
|
|
Db 481 CACTTAGCGTTAAAGCCACTTTTCGGCTATGTTACTATCTTTTAGGTGTAAAC 540
|
|
|
QY 541 CAATCTCTTTTAGCAAGCGCTTACAAAGGAATGATGATGATGATGATGATGATGATGATGATG 600
|
|
|
Db 541 CAATCTCTTTTAGCAAGCGCTTACAAAGGAATGATGATGATGATGATGATGATGATGATG 600
|
|
|
QY 601 CAAGTAGATGGTAAATAAACCAATTCGCCAACAGTGTTCGACGCTATCATTAATTTGAAA 660
|
|
|
Db 601 CAAGTAGATGGTAAATAAACCAATTCGCCAACAGTGTTCGACGCTATCATTAATTTGAAA 660
|
|
|
QY 661 CAATTTAACATCGACCAAAATGAATGGTGCAGTTTTACAAAGAAACAACTCGGCC 720
|
|
|
Db 661 CAATTTAACATCGACCAAAATGAATGGTGCAGTTTTACAAAGAAACAACTCGGCC 720
|
|
|
QY 721 GTATTCAACCGTGTACATCTAACCCTTCCCAATTTAAAGGGATTTTAGATTTCTAAC 780
|
|
|
Db 721 GTATTCAACCGTGTACATCTAACCCTTCCCAATTTAAAGGGATTTTAGATTTCTAAC 780
|
|
|
QY 781 GGACAAAGTCTTTTAAATCAACCCAAATGTTATCACAATAGGTAAAGCAAGCAATTTAAAC 840
|
|
|
Db 781 GGACAAAGTCTTTTAAATCAACCCAAATGTTATCACAATAGGTAAAGCAAGCAATTTAAAC 840
|
|
|
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGGCGCTAAT 900
|
|
|
Db 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAACATCAAGGCGCTAAT 900
|
|
|
QY 901 TTCACCTTCGAGCAAAACCAAGATGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
|
|
|
Db 901 TTCACCTTCGAGCAAAACCAAGATGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
|
|
|
QY 961 ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGCAAGGCTGTG 1020
|
|
|
Db 961 ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGCAAGGCTGTG 1020
|
|
|
QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTTTACTCGAGGGCAAAAATACCATCAGCGAT 1080
|
|
|
Db 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTTTACTCGAGGGCAAAAATACCATCAGCGAT 1080
|
|
|
QY 1081 ATATAAACCACCAACCATTTACTTACAGCATTTGCGGCGCTGAAATGAAGCGCTCAATCTG 1140
|
|
|
Db 1081 ATATAAACCACCAACCATTTACTTACAGCATTTGCGGCGCTGAAATGAAGCGCTCAATCTG 1140
|
|
|
QY 1141 GCGCATATTTTGGCAAAAGCGGTAACTTAATGTCCTGCTGCCACTATTCGAACCA 1200
|
|
|
Db 1141 GCGCATATTTTGGCAAAAGCGGTAACTTAATGTCCTGCTGCCACTATTCGAACCA 1200
|
|
|
QY 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTTGTTCTTTCCGCC 1260
|
|
|
Db 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTTGTTCTTTCCGCC 1260
|
|
|
QY 1261 AAAGAGGTGAAGCGGAAATTTGCGGCTGTAAATTTCCGCTCAAAATCAGCAAGCTTAAGGC 1320
|
|
|
Db 1261 AAAGAGGTGAAGCGGAAATTTGCGGCTGTAAATTTCCGCTCAAAATCAGCAAGCTTAAGGC 1320
|
|
|
QY 1321 GCGAAGCTGATGATTTACAGGCGGATAAAGTCACTATTAATAACAGGTGACGTTATCGACCTT 1380
|
|
|
Db 1321 GCGAAGCTGATGATTTACAGGCGGATAAAGTCACTATTAATAACAGGTGACGTTATCGACCTT 1380
|
|
|

```

QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGAGAGCGCGCGGGAAGGTAAAAAC 1440  
 DB 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGAGAGCGCGCGGGAAGGTAAAAAC 1440  
 QY 1441 GGCATTCAATTAGCAAGAAAGCTCTTTAGAAAGAGGTCAACCATCATGTATCAGCG 1500  
 DB 1441 GGCATTCAATTAGCAAGAAAGCTCTTTAGAAAGAGGTCAACCATCATGTATCAGCG 1500  
 QY 1501 AAAAGAAAAGCGGAGCGCTATTGTGTGGGCGGATATTGCGTTAAATTGACGGCAATATT 1560  
 DB 1501 AAAAGAAAAGCGGAGCGCTATTGTGTGGGCGGATATTGCGTTAAATTGACGGCAATATT 1560  
 QY 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGGTTTTGTGGAGACATCGGGGCAT 1620  
 DB 1561 AACGCTCAAGGTAGTGTGATATCGCTAAAACCGGTGGTTTTGTGGAGACATCGGGGCAT 1620  
 QY 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGGTGTGCTAGACCCCTGAT 1680  
 DB 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGGTGTGCTAGACCCCTGAT 1680  
 QY 1681 GATGTAACAATTGAAGCCGAAGACCCCTTCGCAATTAATACCGGTATTAATGATGATTC 1740  
 DB 1681 GATGTAACAATTGAAGCCGAAGACCCCTTCGCAATTAATACCGGTATTAATGATGATTC 1740  
 QY 1741 CCAACAGGCAACCGGTGAAGCAAGCGCCCTAAAAAATAGCGAATCTCAAAACACGCTA 1800  
 DB 1741 CCAACAGGCAACCGGTGAAGCAAGCGCCCTAAAAAATAGCGAATCTCAAAACACGCTA 1800  
 QY 1801 ACCAATACAACTATTTCAAAATTTATCTGAAAACGCCCTGGACAATGAATATAACGGCATCA 1860  
 DB 1801 ACCAATACAACTATTTCAAAATTTATCTGAAAACGCCCTGGACAATGAATATAACGGCATCA 1860  
 QY 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCTCCAT 1920  
 DB 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCTCCAT 1920  
 QY 1921 AGTAAAGTCAAGCGTGGCGAGGGTTCAGATTGATGGAGATATTACTTCTAAAGGCGGA 1980  
 DB 1921 AGTAAAGTCAAGCGTGGCGAGGGTTCAGATTGATGGAGATATTACTTCTAAAGGCGGA 1980  
 QY 1981 AATTTAACCATTATCTGGCGGATGGTTGATGTTTCATATAAATATATACGCTTGATCAG 2040  
 DB 1981 AATTTAACCATTATCTGGCGGATGGTTGATGTTTCATATAAATATATACGCTTGATCAG 2040  
 QY 2041 GGTTTTTTAATATATACCGCGCTTCGCTAGCTTTTGAAGTGAATTAACAAAGCAGCG 2100  
 DB 2041 GGTTTTTTAATATATACCGCGCTTCGCTAGCTTTTGAAGTGAATTAACAAAGCAGCG 2100  
 QY 2101 GACGGGCAATGCTAAATTTGTCGCCAGGGCACTGTACCAATACAGGAGGGGAAA 2160  
 DB 2101 GACGGGCAATGCTAAATTTGTCGCCAGGGCACTGTACCAATACAGGAGGGGAAA 2160  
 QY 2161 GATTTCAGGGCTAACAGCTATCTTTAAACGGAAACGGGTAAAGTCTGAATATCATTTCA 2220  
 DB 2161 GATTTCAGGGCTAACAGCTATCTTTAAACGGAAACGGGTAAAGTCTGAATATCATTTCA 2220  
 QY 2221 TCAGTGAATAATTTAAACCAAACTTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280  
 DB 2221 TCAGTGAATAATTTAAACCAAACTTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280  
 QY 2281 ATTAACCAAACTACGAGAAAGAACACCTTCGTTATTGGCAACACGCACTGATTCGCACTGG 2340  
 DB 2281 ATTAACCAAACTACGAGAAAGAACACCTTCGTTATTGGCAACACGCACTGATTCGCACTGG 2340  
 QY 2341 AACGTCAGTCTTAACTAGACAGCGCGCAATTTTACCTTTTAAATACATTTCA 2400  
 DB 2341 AACGTCAGTCTTAACTAGACAGCGCGCAATTTTACCTTTTAAATACATTTCA 2400  
 QY 2401 AGCAATAGCAAGGGCTTAACACACAGATTAAGAACCTCTGACGGGGTGAATTTTAAACGGC 2460  
 DB 2401 AGCAATAGCAAGGGCTTAACACACAGATTAAGAACCTCTGACGGGGTGAATTTTAAACGGC 2460

QY 2461 GTAAATGCAACATCTCATTCAATCTCAAGAGAGCGAAAGTTAAATTTCAAAATTTAAA 2520  
 DB 2461 GTAAATGCAACATCTCATTCAATCTCAAGAGAGCGAAAGTTAAATTTCAAAATTTAAA 2520  
 QY 2521 CCAACAGAGACATGAACACAGCAACCTTTACCAATTCGGTTTTTAGCCCAATATCACA 2580  
 DB 2521 CCAACAGAGACATGAACACAGCAACCTTTACCAATTCGGTTTTTAGCCCAATATCACA 2580  
 QY 2581 GCCACTGTGGGGCTCTGTTTTTTTATATATATGCAACCAATCTCTGGCAGAGGGCT 2640  
 DB 2581 GCCACTGTGGGGCTCTGTTTTTTTATATATATGCAACCAATCTCTGGCAGAGGGCT 2640  
 QY 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGCGCTAAATTTTACCTTTAAATTTCCCAT 2700  
 DB 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGCGCTAAATTTTACCTTTAAATTTCCCAT 2700  
 QY 2701 GTTTCGGCGGATGAGCGCTTTTAAAAATCAACAAAGACTTTAAACCATAAATGCAACCAATTTCA 2760  
 DB 2701 GTTTCGGCGGATGAGCGCTTTTAAAAATCAACAAAGACTTTAAACCATAAATGCAACCAATTTCA 2760  
 QY 2761 AATTTACGCTCAGACAGACGAAAGATGATTTTATGAGGGTACGACGCAATGCCATC 2820  
 DB 2761 AATTTACGCTCAGACAGACGAAAGATGATTTTATGAGGGTACGACGCAATGCCATC 2820  
 QY 2821 AATTTACGCTCAGACATATCTGATTTCTGGCGGTAAATGTCACCCCTTGGTGGAACAACTCA 2880  
 DB 2821 AATTTACGCTCAGACATATCTGATTTCTGGCGGTAAATGTCACCCCTTGGTGGAACAACTCA 2880  
 QY 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTGAAGGCC 2940  
 DB 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTGAAGGCC 2940  
 QY 2941 AATAACGCCCTTAATCAGCAAAACATAGGGATAGAGTTATATAAATTTGGCAGCTTGCTC 3000  
 DB 2941 AATAACGCCCTTAATCAGCAAAACATAGGGATAGAGTTATATAAATTTGGCAGCTTGCTC 3000  
 QY 3001 GTTAATGGGAGTTTAAAGTTAACTGGCGAAAATGAGATATTAAAGGCAATCTCACTATT 3060  
 DB 3001 GTTAATGGGAGTTTAAAGTTAACTGGCGAAAATGAGATATTAAAGGCAATCTCACTATT 3060  
 QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATATACCGGCAATTTT 3120  
 DB 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATATACCGGCAATTTT 3120  
 QY 3121 ACCAATTAATGGCATCTCCGAAATTAATATAACACAGAGAGTGTAAACCTTGGCAATGTT 3180  
 DB 3121 ACCAATTAATGGCATCTCCGAAATTAATATAACACAGAGAGTGTAAACCTTGGCAATGTT 3180  
 QY 3181 ACCAATTAATGGCATCTCCGAAATTAATATAACACAGAGAGTGTAAACCTTGGCAATGTT 3240  
 DB 3181 ACCAATTAATGGCATCTCCGAAATTAATATAACACAGAGAGTGTAAACCTTGGCAATGTT 3240  
 QY 3241 GCGGAGATATATCAACAAAAGAGAGCTTTAAATATTATACAGACAGTAAATGATGCT 3300  
 DB 3241 GCGGAGATATATCAACAAAAGAGAGCTTTAAATATTATACAGACAGTAAATGATGCT 3300  
 QY 3301 GAAATCCAAATTTGGCGCAATATCTCGAAAAAGAGGCAACCTACGATTTCTTCCGAT 3360  
 DB 3301 GAAATCCAAATTTGGCGCAATATCTCGAAAAAGAGGCAACCTACGATTTCTTCCGAT 3360  
 QY 3361 AAAATTAATATCAACAAAAGAGTAAACAAATCAAAAAGGGTATTGATGGAGAGGACTTAGT 3420  
 DB 3361 AAAATTAATATCAACAAAAGAGTAAACAAATCAAAAAGGGTATTGATGGAGAGGACTTAGT 3420  
 QY 3421 TCAGATGCGCAAGTAAATGCAACCTTAACCTATTAAAAACCAAGAAATTTGAAATTTGACAGAA 3480  
 DB 3421 TCAGATGCGCAAGTAAATGCAACCTTAACCTATTAAAAACCAAGAAATTTGAAATTTGACAGAA 3480  
 QY 3481 GACCTAGTATTTCAGTTTCAATTAAGCAGATTAACGCCAAAGAGTGTAGAGATTTA 3540  
 DB 3481 GACCTAGTATTTCAGTTTCAATTAAGCAGATTAACGCCAAAGAGTGTAGAGATTTA 3540  
 QY 3541 ACTATTGGCAACAGTAAATGACGGTAAACAGCGGTGCGCGAAGCCCAACAGTAACTTTTAAAC 3600







QY 961 ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTGGTGGCAAGTGAAAAACGAGGGTGTG 1020  
DB 961 ACTGTCGGTAAAGACGCGAGTGTAAATCTTATTGGTGGCAAGTGAAAAACGAGGGTGTG 1020  
QY 1021 ATTAGCCGTAATGGTGGCAGCATTTCTTACTCGCAGGGGCAAAAATCAACCATCAGCGAT 1080  
DB 1021 ATTAGCCGTAATGGTGGCAGCATTTCTTACTCGCAGGGGCAAAAATCAACCATCAGCGAT 1080  
QY 1081 ATAATAAACCCAAACCATTTACTTACAGATTGGCCGCGCTGAAATCAAGCGGTCAATCTG 1140  
DB 1081 ATAATAAACCCAAACCATTTACTTACAGATTGGCCGCGCTGAAATCAAGCGGTCAATCTG 1140  
QY 1141 GCGGATATTTTGGCAAGGCGGTACATTAATGTCCGTGCTGCCACTATTCGAAACCAA 1200  
DB 1141 GCGGATATTTTGGCAAGGCGGTACATTAATGTCCGTGCTGCCACTATTCGAAACCAA 1200  
QY 1201 GGTAAACTTCTGCTGATTTGTAAAGCAAGATAAAGCGGCAATATGTTCTTTCCGCC 1260  
DB 1201 GGTAAACTTCTGCTGATTTGTAAAGCAAGATAAAGCGGCAATATGTTCTTTCCGCC 1260  
QY 1261 AAGAGGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320  
DB 1261 AAGAGGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320  
QY 1321 GCGAAGCTGATGATTACAGCGGATAAGTCACATTTAAACAGGTGCGAGTTATCGACCTT 1380  
DB 1321 GCGAAGCTGATGATTACAGCGGATAAGTCACATTTAAACAGGTGCGAGTTATCGACCTT 1380  
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGGGGCAAGGTAAANAC 1440  
DB 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGGGGCAAGGTAAANAC 1440  
QY 1441 GGCATTCAATTAGCAAGAAACCTTTTAGAAAAGGCTCAACATCAATGTATCAGGC 1500  
DB 1441 GGCATTCAATTAGCAAGAAACCTTTTAGAAAAGGCTCAACATCAATGTATCAGGC 1500  
QY 1501 AAAGAAAAGGGGAGCGGCTATTGTGGGGCGATATTGGGTGTAATTTAGCGGCAATATT 1560  
DB 1501 AAAGAAAAGGGGAGCGGCTATTGTGGGGCGATATTGGGTGTAATTTAGCGGCAATATT 1560  
QY 1561 AACGCTCAAGGTAGTGGTGATATCCCTAAAACCGGTGTTTGTGGAGACATCGGGGCAT 1620  
DB 1561 AACGCTCAAGGTAGTGGTGATATCCCTAAAACCGGTGTTTGTGGAGACATCGGGGCAT 1620  
QY 1621 TATTTATCCATTGACAGCAATGCAATTCCTTAAACAAAGAGTGGTTCCTAGACCCCTGAT 1680  
DB 1621 TATTTATCCATTGACAGCAATGCAATTCCTTAAACAAAGAGTGGTTCCTAGACCCCTGAT 1680  
QY 1681 GATGTAAACAAATTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTATAATGATGAATTC 1740  
DB 1681 GATGTAAACAAATTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTATAATGATGAATTC 1740  
QY 1741 CCAACAGCACCGGTGAAGCAAGCGACCCCTTAAAAAATAGCGAATCAAAACACGCTA 1800  
DB 1741 CCAACAGCACCGGTGAAGCAAGCGACCCCTTAAAAAATAGCGAATCAAAACACGCTA 1800  
QY 1801 ACCAATCAACTATTTCAAAATTTCTGAAAACCGCTTGGACAATGAATATAAGCGCATCA 1860  
DB 1801 ACCAATCAACTATTTCAAAATTTCTGAAAACCGCTTGGACAATGAATATAAGCGCATCA 1860  
QY 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAATCCCACTTAATTTCCCAT 1920  
DB 1861 AGAAAACCTTACCGTTAATAGCTCAATCAATCAACATCGGAAGCAATCCCACTTAATTTCCCAT 1920  
QY 1921 AGTAAAGGTGAGCGGCGGCTTACAGATTGATGGAGATATTACTTCTAAAGCGGA 1980  
DB 1921 AGTAAAGGTGAGCGGCGGCTTACAGATTGATGGAGATATTACTTCTAAAGCGGA 1980  
QY 1981 AATTTAACCAATTTATCTGCGGATGGGTTGATGTTTCATAAAAAATATTACGGTTGATCAG 2040  
DB 1981 AATTTAACCAATTTATCTGCGGATGGGTTGATGTTTCATAAAAAATATTACGGTTGATCAG 2040

QY 2041 GGTTTTTTAAATATTACCGCCGCTTCCGTAGCTTTTGAAGGTGAAATAAACAACGACGC 2100  
DB 2041 GGTTTTTTAAATATTACCGCCGCTTCCGTAGCTTTTGAAGGTGAAATAAACAACGACGC 2100  
QY 2101 GACGCGGCAAAATGCTTAAATTTCTGCCCGCAGGCACGTGTAACCATTTACAGGAGAGGAAAA 2160  
DB 2101 GACGCGGCAAAATGCTTAAATTTCTGCCCGCAGGCACGTGTAACCATTTACAGGAGAGGAAAA 2160  
QY 2161 GATTTTCAGGGCTAACACAGGTATCTTTAAACGGAAGCGGTAAAGGTCTCTGAATATCATTTCA 2220  
DB 2161 GATTTTCAGGGCTAACACAGGTATCTTTAAACGGAAGCGGTAAAGGTCTCTGAATATCATTTCA 2220  
QY 2221 TCAGTGAATATTTAAACCCACAATCTTAGTGGCACAATTTAAACATATCTCGGAATATAACA 2280  
DB 2221 TCAGTGAATATTTAAACCCACAATCTTAGTGGCACAATTTAAACATATCTCGGAATATAACA 2280  
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACGCGCATGATTCGCACCTGG 2340  
DB 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACGCGCATGATTCGCACCTGG 2340  
QY 2341 AACGTCAGTCTCTTAATCTAGACAGAGCGCAAAATTTTACCTTTATTAAATACATTTCA 2400  
DB 2341 AACGTCAGTCTCTTAATCTAGACAGAGCGCAAAATTTTACCTTTATTAAATACATTTCA 2400  
QY 2401 AGCAATAGCAAAAGCTTAAACAAACACAGTATAGAGCTCTGCAGGGGTGAATTTAACGGC 2460  
DB 2401 AGCAATAGCAAAAGCTTAAACAAACACAGTATAGAGCTCTGCAGGGGTGAATTTAACGGC 2460  
QY 2461 GTAATGGCAACATGTCTTAATCTAGACAGAGCGCAAAATTTTACCTTTATTAAATACATTTCA 2520  
DB 2461 GTAATGGCAACATGTCTTAATCTAGACAGAGCGCAAAATTTTACCTTTATTAAATACATTTCA 2520  
QY 2521 CCAAAACGAGAACATGACCAACCAACCTTTACCAATTCGAAAGAGGCAAGTTTAAATTTACCAATTAACA 2580  
DB 2521 CCAAAACGAGAACATGACCAACCAACCTTTACCAATTCGAAAGAGGCAAGTTTAAATTTACCAATTAACA 2580  
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGCGAGAGGGGT 2640  
DB 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCATTTCTGCGAGAGGGGT 2640  
QY 2641 GAGTTTAAATATGATGAAATTAATATCTCTAACGGCGCTTAATTTTACCTTTAAATTTCCCAT 2700  
DB 2641 GAGTTTAAATATGATGAAATTAATATCTCTAACGGCGCTTAATTTTACCTTTAAATTTCCCAT 2700  
QY 2701 GTTCGCGGCGATCAGCTTTTAAATCAACAAGACTTAAACCAATAAATGCAACCAATTTCA 2760  
DB 2701 GTTCGCGGCGATCAGCTTTTAAATCAACAAGACTTAAACCAATAAATGCAACCAATTTCA 2760  
QY 2761 AATTTTCAGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC 2820  
DB 2761 AATTTTCAGCTCAGACAGAGAAAGATGATTTTATGACGGGTACGACGCAATGCCATC 2820  
QY 2821 AATTTAACCTTACAAATATCTCTGCGGCTTAATGTCACCCCTTGGTGGACAAACCTCA 2880  
DB 2821 AATTTAACCTTACAAATATCTCTGCGGCTTAATGTCACCCCTTGGTGGACAAACCTCA 2880  
QY 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940  
DB 2881 AGCAGCAGCATTTACGGGGAATATTACTATCGAGAAAGCAGCAAAATGTTACGCTAGAAGCC 2940  
QY 2941 AATAACGCCCTTAATCAGCAAAACATAGGGATAGATTATATAAATCTGGCAGCTTGCTC 3000  
DB 2941 AATAACGCCCTTAATCAGCAAAACATAGGGATAGATTATATAAATCTGGCAGCTTGCTC 3000  
QY 3001 GTTAATGGAGTTTAAAGTTTAACTGGCGAAATGCAGATATTTAAAGGCAATCTCACTATT 3060  
DB 3001 GTTAATGGAGTTTAAAGTTTAACTGGCGAAATGCAGATATTTAAAGGCAATCTCACTATT 3060  
QY 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACCCGGCAATTTT 3120  
DB 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACCCGGCAATTTT 3120  
QY 3121 ACCAATAATGGCACTGCCGGAATTAATTAACAAGAGGAGTGGTAAAACTTTGGCAATGTT 3180

3121 ACCAATAATGGCACTGCGCAAAATTAATATAACACAAGAGAGTGGTAAACCTTGGCAATGTT 3180  
3181 ACCAATGATGGTATTATTAACATATACCACTACGCTAAACGCAACCAAGAGAGCATCATC 3240  
3181 ACCAATGATGGTATTATTAACATATACCACTACGCTAAACGCAACCAAGAGAGCATCATC 3240  
3241 GCGGAGATATATCAACAAAAAGAGAGCTTAAATATTAACACAGTAAATATATGATGCT 3300  
3241 GCGGAGATATATCAACAAAAAGAGAGCTTAAATATTAACACAGTAAATATATGATGCT 3300  
3301 GAATATCAATATGCGGCAATATCTCGCAAAAGAGAGCAACCTACGATTTCTCCGAT 3360  
3301 GAATATCAATATGCGGCAATATCTCGCAAAAGAGAGCAACCTACGATTTCTCCGAT 3360  
3361 AAAATTAATATCAACCAACAGATTAACAATCAAAAAGGCTATTCATGGAGAGGACTCTAGT 3420  
3361 AAAATTAATATCAACCAACAGATTAACAATCAAAAAGGCTATTCATGGAGAGGACTCTAGT 3420  
3421 TCAGATCGCAACAGTAATGCCAAGCTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480  
3421 TCAGATCGCAACAGTAATGCCAAGCTTAATTAACCAAGAAATTTGAAATTTGACAGAA 3480  
3481 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540  
3481 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGAGATTACAGCCAAAGATGGTAGAGATTTA 3540  
3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCGCAAGGCAACCAAGCAAGTAACTTTAAC 3600  
3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCGCAAGGCAACCAAGCAAGTAACTTTAAC 3600  
3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTCACAAATGTGACACTAAATAGCAAAAGTG 3660  
3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTCACAAATGTGACACTAAATAGCAAAAGTG 3660  
3661 AAAACATCTAGCAAGTGGGAGGCTGAAGCAATAGCGCAACAGTACCGGCTTAAT 3720  
3661 AAAACATCTAGCAAGTGGGAGGCTGAAGCAATAGCGCAACAGTACCGGCTTAAT 3720  
3721 ATTACTGCAAAATGTAGAGTAAACAAAGATATTACTTCTCAAAACAGTAAATATC 3780  
3721 ATTACTGCAAAATGTAGAGTAAACAAAGATATTACTTCTCAAAACAGTAAATATC 3780  
3781 ACCCGCTCGGAAAAGGTTTACCACACAGCAGCGCTGACCATTAACGCAACAAATGGCAAA 3840  
3781 ACCCGCTCGGAAAAGGTTTACCACACAGCAGCGCTGACCATTAACGCAACAAATGGCAAA 3840  
3841 GCAAGTATTACAAACCAACAGGTGATATACGCGGTACGATTTCCGGTAAACACGGTAAAT 3900  
3841 GCAAGTATTACAAACCAACAGGTGATATACGCGGTACGATTTCCGGTAAACACGGTAAAT 3900  
3901 GTTACGCGGACGTGGTATTAACTAACTAAATCCGGCTCAAAAATTTGAAGCAAAATCCGGT 3960  
3901 GTTACGCGGACGTGGTATTAACTAACTAAATCCGGCTCAAAAATTTGAAGCAAAATCCGGT 3960  
3961 GAGGCTAATGTAAAGTGCAACAGTACAAATTTGCGGTGACAAATTTCCGGTAAATACGGTA 4020  
3961 GAGGCTAATGTAAAGTGCAACAGTACAAATTTGCGGTGACAAATTTCCGGTAAATACGGTA 4020  
4021 AATGTTACGGCAACAGCTGGCGATTTACAGTTGGGAATGGCGCAAGAAATTAATCCGACA 4080  
4021 AATGTTACGGCAACAGCTGGCGATTTACAGTTGGGAATGGCGCAAGAAATTAATCCGACA 4080  
4081 GAAGGAGCTGCAACCTTTAACCCCAACAGGGAATACCTTGACTGACTGAAAGCGGTTCTAGC 4140  
4081 GAAGGAGCTGCAACCTTTAACCCCAACAGGGAATACCTTGACTGACTGAAAGCGGTTCTAGC 4140  
4141 ATCACTTTCAACTAAGGCTCAGGTAGACCTCTTTGGCTCAAGATTTGAGTATCGCAGGAAGC 4200  
4141 ATCACTTTCAACTAAGGCTCAGGTAGACCTCTTTGGCTCAAGATTTGAGTATCGCAGGAAGC 4200  
4201 ATTAATGCTGCTAATGTGACATTAATATCTACAGGCACCTTAACCAACCGTGGCAGGCTCG 4260

4201 ATTAATGCTGCTAATGTGACATTAATATCTACAGGCACCTTTAAACCCCGTGGCAGGCTCG 4260  
4261 GATATTAAGCAACACAGCGGCACCTTTGGTATTATTAACGCAAAAGAGTCTTAAGCTAANTGTT 4320  
4261 GATATTAAGCAACACAGCGGCACCTTTGGTATTATTAACGCAAAAGAGTCTTAAGCTAANTGTT 4320  
4321 GATGATCAGGTGATGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTATGTTG 4380  
4321 GATGATCAGGTGATGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTATGTTG 4380  
4381 ACTGGGGCAACCTCAAGCAGTGTGAATATCACTGGGGGATTTTAAACACAGTAAATGGGTTA 4440  
4381 ACTGGGGCAACCTCAAGCAGTGTGAATATCACTGGGGGATTTTAAACACAGTAAATGGGTTA 4440  
4441 AATATCATTTCCAAAAGATGTTAGAAACACTGTGCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500  
4441 AATATCATTTCCAAAAGATGTTAGAAACACTGTGCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500  
4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACCGCTTCTT 4560  
4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGAAACCGCTTCTT 4560  
4561 GAAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAACTTGGTGAAGT 4620  
4561 GAAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAACTTGGTGAAGT 4620  
4621 GCTGTACGTTTTTGTGAGCCAAATAATACTACAGTCAATACAGTCAATACAGTAAATTTTACA 4680  
4621 GCTGTACGTTTTTGTGAGCCAAATAATACTACAGTCAATACAGTCAATACAGTAAATTTTACA 4680  
4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTGTTCTCAAGTGGTAAAT 4740  
4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTGTTCTCAAGTGGTAAAT 4740  
4741 GCGCAGCAGGATGTACCAATGTGCTGACATGAGCGGTGAGCGGTGAGTCAAGTAAATTTGACAAG 4800  
4741 GCGCAGCAGGATGTACCAATGTGCTGACATGAGCGGTGAGCGGTGAGTCAAGTAAATTTGACAAG 4800  
4801 GTAGATTTCACTCTGCAATGAAGTCAATTTTATTTTCTGATTTTACTGTGGTAAAT 4860  
4801 GTAGATTTCACTCTGCAATGAAGTCAATTTTATTTTCTGATTTTACTGTGGTAAAT 4860  
4861 GTTACGATGGGCTTTACCATCTTTTAAATAATTTAGGAGATCAATCAATTAATTTT 4920  
4861 GTTACGATGGGCTTTACCATCTTTTAAATAATTTAGGAGATCAATCAATTAATTTT 4920  
4921 AACAGGTTATTATTATG 4937  
4921 AACAGGTTATTATTATG 4937

RESULT 8  
AR032350  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Score 97.6%;  
98.5%;  
DB 4820.2;  
Length 9323;  
Pred. No. 0;  
i  
Linear PAT 29-SEP-1999  
9323 bp DNA  
6 from patent US 5869065.  
AR032350  
Sequence  
AR032350  
Accession  
AR032350.1  
GI:5947955  
Unknown.  
Unclassified.  
1 (bases 1 to 9323)  
Barenkamp, S. J. and St. Geme, J. William, III.  
High molecular weight surface proteins of non-typeable haemophilus  
Patent: US 5869065-A 6 09-FEB-1999;  
Location/Qualifiers  
source  
3134 a 1829 c 1832 g 2528 t  
/organism="unknown"

Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAAGATATAAAATAAATCAAGATTTTGTGATGACAAACAACAATACAA 60  
Db 441 TAAATATACAAGATATAAAATAAATCAAGATTTTGTGATGACAAACAACAATACAA 500  
QY 61 CACCTTTTTCAGCTATATGCAAAATATTTTAAAAAAATAGTATAAATCCGCCATATA 120  
Db 501 CACCTTTTTCAGCTATATGCAAAATATTTTAAAAAAATAGTATAAATCCGCCATATA 560  
QY 121 AATGGTATATCTTCATCTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTT 180  
Db 561 AATGGTATATCTTCATCTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTT 620  
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680  
QY 241 CACATGAATATGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 681 CACATGAATATGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740  
QY 301 GAACGCAATGATAAGTAATTTAATTTCAACTAACCTTAGGAGAAAAATATGAACAAG 360  
Db 741 GAACGCAATGATAAGTAATTTAATTTCAACTAACCTTAGGAGAAAAATATGAACAAG 800  
QY 361 ATATATCGTCTCAAAATTCAGCAAGCGCTGAATGCTTTGGTGTGCTGCTGAATTTGCA 420  
Db 801 ATATATCGTCTCAAAATTCAGCAAGCGCTGAATGCTTTGGTGTGCTGCTGAATTTGCA 860  
QY 421 CGGGTGTGACCATTCACAGAAAAAGGCTCCGCTATGTACTATCTTTAGGTGTAAC 480  
Db 861 CGGGTGTGACCATTCACAGAAAAAGGCGAAGAACCTGCTGCGCATGAAAGTGGCT 920  
QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTTCCA 540  
Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTTCCA 980  
QY 541 CAATCTGTTTAAAGCGGCTTACAGGAATGATGTAGTACAGGACAGGACACATATG 600  
Db 981 CAATCTGTTTAAAGCGGCAATTTACATCGACCAAAATGAAATGGTGAGTTTATAC 1040  
QY 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTTAATTTGAAA 660  
Db 1041 AAGAAAACAAGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTTAATTTGAAA 1100  
QY 661 CAATTTACATCGACCAAAATGAAATGGTGAGTGTTTTACAGAAAACAACATCCGCC 720  
Db 1101 CAATTTACATCGACCAAAATGAAATGGTGAGTGTTTTACAGAAAACAACATCCGCC 1160  
QY 721 GTATTTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGATTTTATAGTTTAA 780  
Db 1161 GTATTTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGATTTTATAGTTTAA 1220  
QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATACAAATAGTAAAGACGCAATTTAATAC 840  
Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGGTATACAAATAGTAAAGACGCAATTTAATAC 1280  
QY 841 ACTAATGCTTTAGCGCTTCTACCTAGACATTTCTAACGAAACATCAAGGCGGTAAT 900  
Db 1281 ACTAATGCTTTAGCGCTTCTACCTAGACATTTCTAACGAAACATCAAGGCGGTAAT 1340  
QY 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCAGGTTTAAAT 960  
Db 1341 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCAGGTTTAAAT 1400  
QY 961 ACTGTCGGTAAAGACGCGATGTAATCTTTATTTGGTGGCAAGTGAAGGAGGAGGAGG 1020  
Db 1401 ACTGTCGGTAAAGACGCGATGTAATCTTTATTTGGTGGCAAGTGAAGGAGGAGGAGG 1460  
QY 1021 ATTAGCGTAATGGTGACGATTTCTTTTACTCGCAGGCGCAAAATACCATCAGCGAT 1080  
Db 1461 ATTAGCGTAATGGTGACGATTTCTTTTACTCGCAGGCGCAAAATACCATCAGCGAT 1520

QY 1081 ATAAATAACCAACCAATTTACTTTACAGCATTTGCCGCGCTGAAATAGAGCGGTCAATCTG 1140  
Db 1521 ATAAATAACCAACCAATTTACTTTACAGCATTTGCCGCGCTGAAATAGAGCGGTCAATCTG 1580  
QY 1141 GCGATATTTTTCGCAAAAGCGGTAAATTAATGTCGCTGCCACTATTCGAAACCAA 1200  
Db 1581 GCGATATTTTTCGCAAAAGCGGTAAATTAATGTCGCTGCCACTATTCGAAACCAA 1640  
QY 1201 GGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATTAAGCGGCAATATTTCTTTTCGCGC 1260  
Db 1641 GGTAAACTTTCTGCTGATTTCTGTAAAGCAAGATTAAGCGGCAATATTTCTTTTCGCGC 1700  
QY 1261 AAAGAGGTTGAAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320  
Db 1701 AAAGAGGTTGAAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1760  
QY 1321 GGCAAGCTGATGATTACAGCGGATTAAGTACATTTAAAACAGGTGACGTATTCGACCTT 1380  
Db 1761 GGCAAGCTGATGATAAAGTCCGATTAAGTACATTTAAAACAGGTGACGTATTCGACCTT 1820  
QY 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTGGGTGACGAGCGGCGGAGGTAAGAAC 1440  
Db 1821 TCAGGTAAAGAGGGGAGAACTTTACCTTGGGTGACGAGCGGCGGAGGTAAGAAC 1880  
QY 1441 GGCATTCAATTAAGCAAGAAACCTCTTTTAGAAAAAGGCTCAACCATCAATCTATCAGC 1500  
Db 1881 GGCATTCAATTAAGCAAGAAACCTCTTTTAGAAAAAGGCTCAACCATCAATCTATCAGC 1940  
QY 1501 AAAGAAAAAGGCGGAGCGCTATTGTGGGCGCATTTTGGCTTAATTTGACGCAATAT 1560  
Db 1941 AAAGAAAAAGGCGGAGCGCTATTGTGGGCGCATTTTGGCTTAATTTGACGCAATAT 2000  
QY 1561 AACGCTCAAGGTAGTGTGATATCCGTAAACCGGTGTTTTGTGGAGACATCGGGCAT 1620  
Db 2001 AACGCTCAAGGTAGTGTGATATCCGTAAACCGGTGTTTTGTGGAGACATCGGGCAT 2060  
QY 1621 TATTATCCATTTGACGCAATGCAATTTGTTAAACAAAAAGAGTGTGTAGACCTGAT 1680  
Db 2061 TATTATCCATTTGACGCAATGCAATTTGTTAAACAAAAAGAGTGTGTAGACCTGAT 2120  
QY 1681 GATGTAAACATTTGACGCAAGACCCCTTCGCAATATACCGGTATTAATGATGAATTC 1740  
Db 2121 GATGTAAACATTTGACGCAAGACCCCTTCGCAATATACCGGTATTAATGATGAATTC 2180  
QY 1741 CCAACAGCAGCGGTGAAAGCAAGACCCCTTAAACAAAAATAGGAACTCAAAACACGCTA 1800  
Db 2181 CCAACAGCAGCGGTGAAAGCAAGACCCCTTAAACAAAAATAGGAACTCAAAACACGCTA 2240  
QY 1801 ACCAATACACTATTTTCAAAATTTCTGAAAACGCGCTGGACAATGAATATAACGCGATCA 1860  
Db 2241 ACCAATACACTATTTTCAAAATTTCTGAAAACGCGCTGGACAATGAATATAACGCGATCA 2300  
QY 1861 AGAAACTTACCGTTTAAATAGCTCAATCAACATCGGAGCAACTCCCACTTAATTTCTCCAT 1920  
Db 2301 AGAAACTTACCGTTTAAATAGCTCAATCAACATCGGAGCAACTCCCACTTAATTTCTCCAT 2360  
QY 1921 AGTAAAGGTACGCTGCGGAGCGGCTTCAGATTGATGAGATATTTACTTTCTAAAGCGGGA 1980  
Db 2361 AGTAAAGGTACGCTGCGGAGCGGCTTCAGATTGATGAGATATTTACTTTCTAAAGCGGGA 2420  
QY 1981 AATTAAACCATTTTATTTGCGGCGATGGTTCATATAAATAATTTACGCTTGATCAG 2040  
Db 2421 AATTAAACCATTTTATTTGCGGCGATGGTTCATATAAATAATTTACGCTTGATCAG 2480  
QY 2041 GGTTTTTTAAATATTTACGCGCTTCCGTAGCTTTTGAAGGTGGAATAACAAAGCAGC 2100  
Db 2481 GGTTTTTTAAATATTTACGCGCTTCCGTAGCTTTTGAAGGTGGAATAACAAAGCAGC 2540  
QY 2101 GACCGGCAATGCTAAAAATTTGTCGCCAGGCGCTGTAACTTACAGGAGAGGAAAA 2160  
Db 2541 GACCGGCAATGCTAAAAATTTGTCGCCAGGCGCTGTAACTTACAGGAGAGGAAAA 2600





Db 4761 GATGATCAGGTGATAGTACAGAAGTGAAATGCAGTCAACGACTGGGATTTGGTAGTGTG 4820  
QY 4381 ACTGGCGCAACCTCAAGCAGTGTGAATATCACTACCTGGGGATTTAAACACAGTAATGGGTTA 4440  
Db 4821 ACTGGCGCAACCTCAAGCAGTGTGAATATCACTACCTGGGGATTTAAACACAGTAATGGGTTA 4880  
QY 4441 AATATCATTTTCAAGAGATGGTGAAGAACACTGTGCGCTTTAAGAGGCAAGAAATGGAGTGTG 4500  
Db 4881 AATATCATTTTCAAGAGATGGTGAAGAACACTGTGCGCTTTAAGAGGCAAGAAATGGAGTGTG 4940  
QY 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAATTTGAAGCGGAAAGCGCTCTT 4560  
Db 4941 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAATTTGAAGCGGAAAGCGCTCTT 5000  
QY 4561 GAAAAGTAAAGATTTATCTGATGAAGAGAGAGAAACATTAGCTTAAACTTTGGTGTAAAGT 4620  
Db 5001 GAAAAGTAAAGATTTATCTGATGAAGAGAGAGAAACATTAGCTTAAACTTTGGTGTAAAGT 5060  
QY 4621 GCTGTACGTTTGTGAGCCAAATTAATCAATTTAGAGTCAATACAGTCAATACAAATGAATTTACA 4680  
Db 5061 GCTGTACGTTTGTGAGCCAAATTAATCAATTTAGAGTCAATACAGTCAATACAAATGAATTTACA 5120  
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAAT 4740  
Db 5121 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAAT 5180  
QY 4741 GCGCACGAGTATGACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACAAAG 4800  
Db 5181 GCGCACGAGTATGACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACAAAG 5240  
QY 4801 GTAGATTTTCATCCCTCAATGAAGTCAATTTATTTTCGTTATTTTACGTTGTTGGTAAA 4860  
Db 5241 GTAGATTTTCATCCCTCAATGAAGTCAATTTATTTTCGTTATTTTACGTTGTTGGTAAA 5300  
QY 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAAGAAATTTACGAGAAATACAAATAAAGTATTTT 4920  
Db 5301 GTTCAGTACGGGCTTTACCCATCTTTGTAAGAAATTTACGAGAAATACAAATAAAGTATTTT 5360  
QY 4921 AACAGGTTATTATTATG 4937  
Db 5361 AACAGGTTATTATTATG 5377

RESULT 9  
AR083957  
LOCUS AR083957 9323 bp DNA linear PAT 01-SEP-2000  
DEFINITION Sequence 6 from patent US 5977336.  
ACCESSION AR083957  
VERSION AR083957.1 GI:10010728  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 9323)  
AUTHORS Barenkamp, S.J.  
TITLE High molecular weight surface proteins of non-typeable haemophilus  
JOURNAL Patent: US 5977336-A 6 02-NOV-1999;  
FEATURES  
Location/Qualifiers  
source  
1..9323  
/organism="unknown"  
BASE COUNT 3134 a 1829 c 1832 g 2528 t  
ORIGIN

Query Match 97.6%; Score 4820.2; DB 6; Length 9323;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGATAATAAATAAATCAAGATTTTGTGATGACAAACAAATTTACA 60  
Db 441 TAAATATACAGATAATAAATAAATCAAGATTTTGTGATGACAAACAAATTTACA 500  
QY 61 CACCTTTTTCAGCTGTATATGCAATATTTTAAAAAATAGTAAATCGCCATATAA 120  
|||||

Db 501 CACCTTTTTCAGCTGTATATGCAATATTTTAAAAAATAGTAAATCGCCATATAA 560  
QY 121 AATGTTAATCTTTTCATCTTTTCATCTTTAATCTTTTCATCTTTTCATCTTTTCAT 180  
|||||  
Db 561 AATGTTAATCTTTTCATCTTTTCATCTTTTCACTTTTCATCTTTTCATCTTTTCAT 620  
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 240  
|||||  
Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 680  
QY 241 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 681 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740  
QY 301 GAACCAATATGATGAATTAATTTAATTTGTTCAACTTAACCTTAGGAGAAAATATGAACAAG 360  
Db 741 GAACCAATATGATGAATTAATTTAATTTGTTCAACTTAACCTTAGGAGAAAATATGAACAAG 800  
QY 361 ATATATCTGCTCAATTTACGAAACGCCCTGAATGCTTTGTTGCTGTCTGCTGAATGGCA 420  
Db 801 ATATATCTGCTCAATTTACGAAACGCCCTGAATGCTTTGTTGCTGTCTGCTGAATGGCA 860  
QY 421 CGGGCTTTGACCATTTCCACAGAAAAGGCTTCCGCTATGTTACTATCTTTAGGTGTAA 480  
Db 861 CGGGCTTTGACCATTTCCACAGAAAAGGCTTCCGCTATGTTACTATCTTTAGGTGTAA 920  
QY 481 CACTTAGGCTTTAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACTATTTC 540  
Db 921 CACTTAGGCTTTAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACTATTTC 980  
QY 541 CAATCTGTTTAAAGCGGCTTTACAGGAATGGATGTAGTACACGGCAGACCCACTATG 600  
Db 981 CAATCTGTTTAAAGCGGCAATTTAATCATCGACCAAAATGAAATGCTGCTGCTGCTTAC 1040  
QY 601 CAAGTAGATGTAATAAAGCAATTTCCGCAACAGCTTTGACGCTATCAATTAATGGAAA 660  
Db 1041 AAGAAAAGTAATAAAGCAATTTCCGCAACAGCTTTGACGCTATCAATTAATGGAAA 1100  
QY 661 CAATTTAATACGACCAAAATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Db 1101 CAATTTAATACGACCAAAATGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1160  
QY 721 GTATTAACCGCTTTACATCTTAACCAATCTCCCAATTTAAAGGATTTTAGATTTTAA 780  
Db 1161 GTATTAACCGCTTTACATCTTAACCAATCTCCCAATTTAAAGGATTTTAGATTTTAA 1220  
QY 781 GGACAAGTCTTTTAAATCAACCAATGATACCAATGATGATGATGATGATGATGATGAT 840  
Db 1221 GGACAAGTCTTTTAAATCAACCAATGATGATGATGATGATGATGATGATGATGAT 1280  
QY 841 ACTAATGCTTTTACGGCTTTACGCTAGACATTTCTTAACGAAAACATCAAGCGGCTAAT 900  
Db 1281 ACTAATGCTTTTACGGCTTTACGCTAGACATTTCTTAACGAAAACATCAAGCGGCTAAT 1340  
QY 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 1341 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1400  
QY 961 ACTGTCGGTAAAGAGCGGCTGTAATCTTATTTGCTGCAAAAGTCAAAACAGGCTGTG 1020  
Db 1401 ACTGTCGGTAAAGAGCGGCTGTAATCTTATTTGCTGCAAAAGTCAAAACAGGCTGTG 1460  
QY 1021 ATTACGCTAAATGCTGCGAGCATTTCTTTTACTCGCAGGCAAAAATCACCATCAGCAT 1080  
Db 1461 ATTACGCTAAATGCTGCGAGCATTTCTTTTACTCGCAGGCAAAAATCACCATCAGCAT 1520  
QY 1081 ATAATAACCAACCATTTACTTACAGCATTTGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140  
Db 1521 ATAATAACCAACCATTTACTTACAGCATTTGCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1580  
QY 1141 GCGCATATTTTGGCAAGCGGTAACATTAATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
Db 1581 GCGCATATTTTGGCAAGCGGTAACATTAATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1640

Qy	1201	GGTAACCTTCTGCTGATTTCTGTAAGCAAGATAAAAGCGCAATATTTGTTCTTCCGCC	1260
Db	1641	GGTAACCTTCTGCTGATTTCTGTAAGCAAGATAAAAGCGCAATATTTGTTCTTCCGCC	1700
Qy	1261	AAAGAGGTTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC	1320
Db	1701	AAAGAGGTTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC	1760
Qy	1321	GGCAAGCTGATGATTTACAGCGGTAAGTCAATTTAAACAGAGTGCAGTTATTCACGCTT	1380
Db	1761	GGCAAGCTGATGATTTAAAGTCCGATTAAGTCAATTTAAACAGAGTGCAGTTATTCACGCTT	1820
Qy	1381	TCAGGTTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGGTAAAC	1440
Db	1821	TCAGGTTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGGTAAAC	1880
Qy	1441	GGCAATTCATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC	1500
Db	1881	GGCAATTCATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC	1940
Qy	1501	AAAGAAAAGCGGACGCGCTATTTGTGGGGCGGATATTGCGTTAATTGACGGCAATATT	1560
Db	1941	AAAGAAAAGCGGACGCGCTATTTGTGGGGCGGATATTGCGTTAATTGACGGCAATATT	2000
Qy	1561	AACGCTCAAGGTAGTGGTGATATCGCTTAAACACCGGTGGTTTGTGGAGACATCGGGGCAT	1620
Db	2001	AACGCTCAAGGTAGTGGTGATATCGCTTAAACACCGGTGGTTTGTGGAGACATCGGGGCAT	2060
Qy	1621	TATTTATCCATTGACAGCAATGCAATTTGTTAAACACCGGTGGTTTGTGGAGACATCGGGGCAT	1680
Db	2061	TATTTATCCATTGACAGCAATGCAATTTGTTAAACACCGGTGGTTTGTGGAGACATCGGGGCAT	2120
Qy	1681	GATGTAAACATTTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTGATTAATGATGAATTC	1740
Db	2121	GATGTAAACATTTGAAGCGGAAGACCCCTTCGCAATTAATACCGGTGATTAATGATGAATTC	2180
Qy	1741	CCACAGCGCGGTGAAGCAAGCGCCCTTAAACACCGGTGGTTTGTGGAGACATCGGGGCAT	1800
Db	2181	CCACAGCGCGGTGAAGCAAGCGCCCTTAAACACCGGTGGTTTGTGGAGACATCGGGGCAT	2240
Qy	1801	ACCAATACAACTATTTCAATTTATCTGAACAGCGCTGACCAATGAATATTAACGGCATCA	1860
Db	2241	ACCAATACAACTATTTCAATTTATCTGAACAGCGCTGACCAATGAATATTAACGGCATCA	2300
Qy	1861	AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATCTCCAT	1920
Db	2301	AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATCTCCAT	2360
Qy	1921	AGTAAAGGTACGCGTGGCGGAGCGCTTCAGATTGATGGAGATATTACTTCTAAAGCGGA	1980
Db	2361	AGTAAAGGTACGCGTGGCGGAGCGCTTCAGATTGATGGAGATATTACTTCTAAAGCGGA	2420
Qy	1981	AATTTAACCAATTTATTTCTGGCGGATGGTTGATGTTCAATAAAATATTACGCTTGATCAG	2040
Db	2421	AATTTAACCAATTTATTTCTGGCGGATGGTTGATGTTCAATAAAATATTACGCTTGATCAG	2480
Qy	2041	GGTTTTTAATTTATACCGCGCTTCCGTAGCTTTTGAAGGTGGAAATTAACAAAGCACGC	2100
Db	2481	GGTTTTTAATTTATACCGCGCTTCCGTAGCTTTTGAAGGTGGAAATTAACAAAGCACGC	2540
Qy	2101	GACGCGCAATGCTAAATTTGCGCCAGGCGCTTAAACACCGGTGAAAGTCTGAATATCAATTC	2160
Db	2541	GACGCGCAATGCTAAATTTGCGCCAGGCGCTTAAACACCGGTGAAAGTCTGAATATCAATTC	2600
Qy	2161	GATTTACGGCTTAACAGCTATCTTTAAACGGAACGGGTAAAGTCTGAATATCAATTC	2220
Db	2601	GATTTACGGCTTAACAGCTATCTTTAAACGGAACGGGTAAAGTCTGAATATCAATTC	2660
Qy	2221	TCAGTGAATATTTAAACCGCAATCTTTAGTGGCAGCAATTAACATATCTGGAATATAACA	2280
Db	2661	TCAGTGAATATTTAAACCGCAATCTTTAGTGGCAGCAATTAACATATCTGGAATATAACA	2720
Qy	2281	ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAAAACACGACCATGATTCGCACTGG	2340
Db	2721	ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAAAACACGACCATGATTCGCACTGG	2780
Qy	2341	AACGTGAGTCTCTTAATCTAGACAGAGCGCAATTTTACCTTTTATTAATATACATTTCA	2400
Db	2781	AACGTGAGTCTCTTAATCTAGACAGAGCGCAATTTTACCTTTTATTAATATACATTTCA	2840
Qy	2401	AGCAATAGCAAAAGGCTTAAACACACAGTATAGAACTCTCGAGGGTGAATTTTAAACGGC	2460
Db	2841	AGCAATAGCAAAAGGCTTAAACACACAGTATAGAACTCTCGAGGGTGAATTTTAAACGGC	2900
Qy	2461	GTAATAGCAAAAGGCTTAAACACACAGTATAGAACTCTCGAGGGTGAATTTTAAACGGC	2520
Db	2901	GTAATAGCAAAAGGCTTAAACACACAGTATAGAACTCTCGAGGGTGAATTTTAAACGGC	2960
Qy	2521	CCAAACGAGAACATGAACACACAAAGCAACCTTTTACCAATTCGGTTTATAGCAATATCA	2580
Db	2961	CCAAACGAGAACATGAACACACAAAGCAACCTTTTACCAATTCGGTTTATAGCAATATCA	3020
Qy	2581	GCCACTGGTGGGGGCTCTGTTTTTGTATATATATGCAACCACTCTGCGAGAGGGCT	2640
Db	3021	GCCACTGGTGGGGGCTCTGTTTTTGTATATATATGCAACCACTCTGCGAGAGGGCT	3080
Qy	2641	GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGCTTAATTTTACCTTAAATTTCCCAT	2700
Db	3081	GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGCTTAATTTTACCTTAAATTTCCCAT	3140
Qy	2701	GTTCCGGCGGATGACGCTTTTAAATAATCAACAAAGACCTTAACCATTAATGCAACCAATTC	2760
Db	3141	GTTCCGGCGGATGACGCTTTTAAATAATCAACAAAGACCTTAACCATTAATGCAACCAATTC	3200
Qy	2761	AATTTTACGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	2820
Db	3201	AATTTTACGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	3260
Qy	2821	AATTTTACGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	3280
Db	3261	AATTTTACGCTCAGACAGAGCAAGATGATTTTATGAGGGGTACGACGCAATGCGCATC	3320
Qy	2881	AGCAGCAGCATATGCGGGAATTTACTATCGAGAAAGCAGCAATGTTAGCTAGAGGCC	2940
Db	3321	AGCAGCAGCATATGCGGGAATTTACTATCGAGAAAGCAGCAATGTTAGCTAGAGGCC	3380
Qy	2941	AATAACCCCTTAATCAGCAAAACATAAGGATAGATTATTAACCTTGGCAGCTTGCTC	3000
Db	3381	AATAACCCCTTAATCAGCAAAACATAAGGATAGATTATTAACCTTGGCAGCTTGCTC	3440
Qy	3001	GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTAAGGCAATCTCACTATT	3060
Db	3441	GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATTCAGATATTAAGGCAATCTCACTATT	3500
Qy	3061	TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCTTAAATATATACCGGCAATTTT	3120
Db	3501	TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCTTAAATATATACCGGCAATTTT	3560
Qy	3121	ACCAATATGCACTGCGCAATTAATAACACAGAGGAGTGGTAAACCTTTGGCAATGTT	3180
Db	3561	ACCAATATGCACTGCGCAATTAATAACACAGAGGAGTGGTAAACCTTTGGCAATGTT	3620
Qy	3181	ACCAATATGCACTGCGCAATTAATAACACAGAGGAGTGGTAAACCTTTGGCAATGTT	3240
Db	3621	ACCAATATGCACTGCGCAATTAATAACACAGAGGAGTGGTAAACCTTTGGCAATGTT	3680
Qy	3241	GCGGAGATATTAATCAACAAAAGAGAGCTTAAATATTTACAGACAGTAATATGATGCT	3300
Db	3681	GCGGAGATATTAATCAACAAAAGAGAGCTTAAATATTTACAGACAGTAATATGATGCT	3740
Qy	3301	GAAATCCAAATTTGGGGCAATATCTCGCAAAAAGAGGCACTCACGATTTCTCCGAT	3360
Db	3741	GAAATCCAAATTTGGGGCAATATCTCGCAAAAAGAGGCACTCACGATTTCTCCGAT	3800
Qy	3361	AAAATTAATATACCAACCAAGAGATACAAATCAAAAAGGGTATTGATGGAGGAGGACTAGT	3420





Db	621	CTTTCATCTTTCACTCTTTCACTCTTTCACTCTTTCACTCTTTCACTCTTTCACTCTTTCACTCTTT	680
Qy	241	CACATGAATGATGAACCGAGGGAAGGAGGAGGGAAGAAATGAAGAGGAGCTGAAC	300
Db	681	CACATGAATGATGAACCGAGGGAAGGAGGAGGGAAGAAATGAAGAGGAGCTGAAC	740
Qy	301	GAACGCAATGATAAAGATAATTAATTTGTTCAACTAACCTTAGAGAGAAATATGAACAAG	360
Db	741	GAACGCAATGATAAAGATAATTAATTTGTTCAACTAACCTTAGAGAGAAATATGAACAAG	800
Qy	361	ATATATCGTCTCAAAATTCAGCAAAACGCCCTGAATGCTTTGGTTGCTGTCTGTAATTGGCA	420
Db	801	ATATATCGTCTCAAAATTCAGCAAAACGCCCTGAATGCTTTGGTTGCTGTCTGTAATTGGCA	860
Qy	421	CGGGTGTGTGACCATTTCCACAGAAAAAGGCTTCGGCTATGTTACTATCTTTAGTGTGAAC	480
Db	861	CGGGTGTGTGACCATTTCCACAGAAAAAGGCTTCGGCTATGTTACTATCTTTAGTGTGAAC	920
Qy	481	CACTTAGCGTTAAAGCCACATTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTGCCA	540
Db	921	CACTTAGCGTTAAAGCCACATTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTGCCA	980
Qy	541	CAATCTCTTTTAGCAAGCGGCTTTACAAGGAATGGATGTAGTACACGACACACGCATCTATG	600
Db	981	CAATCTCTTTTAGCAAGCGGCAATTTAAACATCGACCAAAATCAAAATGGTGCAGTTTTCAC	1040
Qy	601	CAAGCTAGATGTAATAAAACCATTAATCCGCACAGCTGTGAGGCTATCTATTAATTGGAAA	660
Db	1041	AAGAAAACAAGTAATAAACCATTAATCCGCACAGCTGTGAGGCTATCTATTAATTGGAAA	1100
Qy	661	CAATTTAAACATCAGCAAAATGAATGGTGAGTTTTTACAAAGAAAACAACACTCGGCC	720
Db	1101	CAATTTAAACATCAGCAAAATGAATGGTGAGTTTTTACAAAGAAAACAACACTCGGCC	1160
Qy	721	GTATTTCAACCGTGTACATCTACCAAAATCTCCCAATTTAAAGGGATTTTAGATTTCTAAC	780
Db	1161	GTATTTCAACCGTGTACATCTACCAAAATCTCCCAATTTAAAGGGATTTTAGATTTCTAAC	1220
Qy	781	GGACAGCTCTTTTAATCAACCCCAATGGTATCAATAGTAAAGACGCGCAATTTATTAAAC	840
Db	1221	GGACAGCTCTTTTAATCAACCCCAATGGTATCAATAGTAAAGACGCGCAATTTATTAAAC	1280
Qy	841	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTTCAACGAAAAACATCAAGGCGCGTAAT	900
Db	1281	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTTCAACGAAAAACATCAAGGCGCGTAAT	1340
Qy	901	TTCACTCTCGAGCAACACCAAGATAAAGCGTTCGCTGAAATTTGTGAATCACGGTTAAATTT	960
Db	1341	TTCACTCTCGAGCAACACCAAGATAAAGCGTTCGCTGAAATTTGTGAATCACGGTTAAATTT	1400
Qy	961	ACTGTCGGTAAAGACGGCAGTCTAAATCTTATTTGGTGCAAGTGAAGAAAACGAGGGTGTG	1020
Db	1401	ACTGTCGGTAAAGACGGCAGTCTAAATCTTATTTGGTGCAAGTGAAGAAAACGAGGGTGTG	1460
Qy	1021	ATTAGCGTAAATTTGGTGGCAGCATTTCTTTACTTCGACGGCAAAAAATTCACCATCACGGAT	1080
Db	1461	ATTAGCGTAAATTTGGTGGCAGCATTTCTTTACTTCGACGGCAAAAAATTCACCATCACGGAT	1520
Qy	1081	ATAATAAACCCCAACCATTTACTTTACAGCATTCGCGGCTGGAATAGAGGGTCAATCTG	1140
Db	1521	ATAATAAACCCCAACCATTTACTTTACAGCATTCGCGGCTGGAATAGAGGGTCAATCTG	1580
Qy	1141	GGCGATATTTTGGCCAAAGGCGGTAAACATTAATCTCGCTGCCACTATTGGAACACAA	1200
Db	1581	GGCGATATTTTGGCCAAAGGCGGTAAACATTAATCTCGCTGCCACTATTGGAACACAA	1640
Qy	1201	GGTAAACTTTCGCTGATCTGTGAAGCAAGATAAAGCGGCAATATGTTCTTTCCGCC	1260
Db	1641	GGTAAACTTTCGCTGATCTGTGAAGCAAGATAAAGCGGCAATATGTTCTTTCCGCC	1700
Qy	1261	AAAGAGGTGAAGCGGAAATTTGGCGGTGTAAATTTCCGCTCAAAATCAGCAAGCTAAAGC	1320
Db	1701	AAAGAGGTGAAGCGGAAATTTGGCGGTGTAAATTTCCGCTCAAAATCAGCAAGCTAAAGC	1760

Qy	1321	GGCAAGCTGATGATTACAGCGCATAAAGTCACATTTAAAAACAGGTGCAGTATTCGACCTT	1380
Db	1761	GGCAAGCTGATGATAAAGTCCGATAAAGTCACAATTTAAAAACAGGTGCAGTATTCGACCTT	1820
Qy	1381	TCAGGTAAAGAAGGGGAGAAAACCTTACCCTTGCGCGTGACGAGCGCGCGGAAGGTAAAAAC	1440
Db	1821	TCAGGTAAAGAAGGGGAGAAAACCTTACCCTTGCGCGTGACGAGCGCGCGGAAGGTAAAAAC	1880
Qy	1441	GSCATTCAAATTAGCAAGAAGAACCTCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC	1500
Db	1881	GSCATTCAAATTAGCAAGAAGAACCTCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC	1940
Qy	1501	AAAGAAAAAGCGCAGCGCATATCTGTGGGGCGGATATTCGCTTAATTTGACGGCAATATT	1560
Db	1941	AAAGAAAAAGCGCAGCGCATATCTGTGGGGCGGATATTCGCTTAATTTGACGGCAATATT	2000
Qy	1561	RACGCTCAAGGTAGTGGTGATATCCGTAAACCGCGTGTTTTGTGGAGACATPCGGGGCAT	1620
Db	2001	RACGCTCAAGGTAGTGGTGATATCCGTAAACCGCGTGTTTTGTGGAGACATPCGGGGCAT	2060
Qy	1621	TATTTATCCATTGACAGCAATGCAATGTGTAAACAAAAGAGTGGTTGCTAGACCCCTGAT	1680
Db	2061	TATTTATCCATTGACAGCAATGCAATGTGTAAACAAAAGAGTGGTTGCTAGACCCCTGAT	2120
Qy	1681	GATGTAAACAANTTGAACCGCGAAGACCCCCTTCGCAATATACCGGTATAAATGATCAAA TTC	1740
Db	2121	GATGTAAACAANTTGAACCGCGAAGACCCCCTTCGCAATATACCGGTATAAATGATCAAA TTC	2180
Qy	1741	CCAACAGCACCGGTGAAGCAAGCGACCCCTTAAAAAATACCGACTCAAACACGCCTA	1800
Db	2181	CCAACAGCACCGGTGAAGCAAGCGACCCCTTAAAAAATACCGACTCAAACACGCCTA	2240
Qy	1801	ACCAATACAACATTTTCAAAATATCTGAAAAACGCCCTGGACAATGAATATAACGGCATCA	1860
Db	2241	ACCAATACAACATTTTCAAAATATCTGAAAAACGCCCTGGACAATGAATATAACGGCATCA	2300
Qy	1861	AGAAACATTACGGTTAATAGCTCAATCAACATPCGGAAGCAACTCCCACCTTAATTCCTCAT	1920
Db	2301	AGAAACATTACGGTTAATAGCTCAATCAACATPCGGAAGCAACTCCCACCTTAATTCCTCAT	2360
Qy	1921	AGTAAAGTCAGCGTGGCGGAGGGCTTCAGATTTGATGGAGATATATCTCTTAAAGGCGGA	1980
Db	2361	AGTAAAGTCAGCGTGGCGGAGGGCTTCAGATTTGATGGAGATATATCTCTTAAAGGCGGA	2420
Qy	1981	AATTTAACCATTTATCTTGGCGGATGGGTGATGTTTCATAAAAAATTTACCTTCGATCAG	2040
Db	2421	AATTTAACCATTTATCTTGGCGGATGGGTGATGTTTCATAAAAAATTTACCTTCGATCAG	2480
Qy	2041	GGTTTTTTTAAATATTACCGCGCTTCCGTACCTTTTGAAGCTGGAAATAACAAACGACGC	2100
Db	2481	GGTTTTTTTAAATATTACCGCGCTTCCGTACCTTTTGAAGCTGGAAATAACAAACGACGC	2540
Qy	2101	GACCGGCAATGCTAAATTTGTCGCCAGGGCACTGTAAACCATTTACAGGAGAGGGAAAA	2160
Db	2541	GACCGGCAATGCTAAATTTGTCGCCAGGGCACTGTAAACCATTTACAGGAGAGGGAAAA	2600
Qy	2161	GATTTCAGGGCTACACAGTATCTTTTAAACGGAAACGGGTAAAGGTCTGAAATATCATTTCA	2220
Db	2601	GATTTCAGGGCTACACAGTATCTTTTAAACGGAAACGGGTAAAGGTCTGAAATATCATTTCA	2660
Qy	2221	TCAGTGAATTAATTTAACCCACAATCTTATAGTGGCAACAATTAACATATCTGGGAATATAACA	2280
Db	2661	TCAGTGAATTAATTTAACCCACAATCTTATAGTGGCAACAATTAACATATCTGGGAATATAACA	2720
Qy	2281	ATTAACCAAACTACGAGAAAGAACCCCTGCTATTTGGCAAAACCCAGCCATGATTCGCACTGG	2340
Db	2721	ATTAACCAAACTACGAGAAAGAACCCCTGCTATTTGGCAAAACCCAGCCATGATTCGCACTGG	2780
Qy	2341	AACGTCAGTGCCTTATCTATCTAGAGACAGGCGCAAAATTTTACCTTTTATTAATACATTTCA	2400
Db	2781	AACGTCAGTGCCTTATCTATCTAGAGACAGGCGCAAAATTTTACCTTTTATTAATACATTTCA	2840

QY 2401 AGCAATAGCAAGGCTTAAACACACAGATATAGAAGCTCTCGAGGGTGAATTTTAAACGCC 2460  
DB 2841 AGCAATAGCAAGGCTTAAACACACAGATATAGAAGCTCTCGAGGGTGAATTTTAAACGCC 2900  
QY 2461 GTAATGGCAACATGTCATTCATCTCAAGAGAGGAGGAAAGTTAAATTTCAATTAATAA 2520  
DB 2901 GTAATGGCAACATGTCATTCATCTCAAGAGAGGAGGAAAGTTAAATTTCAATTAATAA 2960  
QY 2521 CCAACGAGAACATCAACACAGCAAACTTTACCAGTTTCGGTTTAAAGCCAAATATACA 2580  
DB 2961 CCAACGAGAACATCAACACAGCAAACTTTACCAGTTTCGGTTTAAAGCCAAATATACA 3020  
QY 2581 GCCACTGGTGGGCTCTGTTTTTTTGGATATATATGCAACCAATCTCGCAGAGGGCT 2640  
DB 3021 GCCACTGGTGGGCTCTGTTTTTTTGGATATATATGCAACCAATCTCGCAGAGGGCT 3080  
QY 2641 GAGTTAAATAGAGTGAATTAATATCTACAGCGGCTAATTTTACCTTAAATTTCCCAT 2700  
DB 3081 GAGTTAAATAGAGTGAATTAATATCTACAGCGGCTAATTTTACCTTAAATTTCCCAT 3140  
QY 2701 GTTCGGCGGATGAGCTTTTAAATCAACAAAGACTTAAACCATAAATGCAACCAATTTCA 2760  
DB 3141 GTTCGGCGGATGAGCTTTTAAATCAACAAAGACTTAAACCATAAATGCAACCAATTTCA 3200  
QY 2761 AATTTACGCTCAGACAGCAAGAGATGTTTTATGAGGGTACGACAGCAATGCCATC 2820  
DB 3201 AATTTACGCTCAGACAGCAAGAGATGTTTTATGAGGGTACGACAGCAATGCCATC 3260  
QY 2821 AATTTCAACCTACACATATCCATCTGGGCGTATATGTCACCTTTGGTGACAAACTCA 2880  
DB 3261 AATTTCAACCTACACATATCCATCTGGGCGTATATGTCACCTTTGGTGACAAACTCA 3320  
QY 2881 AGCAGCAGATTTACGGGGAATTTACTATCGAGAAAGCAGCAAAATGTTACGGCTAGAAGCC 2940  
DB 3321 AGCAGCAGATTTACGGGGAATTTACTATCGAGAAAGCAGCAAAATGTTACGGCTAGAAGCC 3380  
QY 2941 AATAACGCCCTAATCAGCAAAACATAAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000  
DB 3381 AATAACGCCCTAATCAGCAAAACATAAGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3440  
QY 3001 GTTAATGGAGTTTAACTGGCGAATATGAGATATTTAAAGGCAATCTCACTATT 3060  
DB 3441 GTTAATGGAGTTTAACTGGCGAATATGAGATATTTAAAGGCAATCTCACTATT 3500  
QY 3061 TCAGAAAGGCCACTTTTAAAGGAAGACTAGAGATACCTTAATATACACGGCAATTTT 3120  
DB 3501 TCAGAAAGGCCACTTTTAAAGGAAGACTAGAGATACCTTAATATACACGGCAATTTT 3560  
QY 3121 ACCAATATGGCACTGCCGAAATTAATATACCAAGGAGTGGTAAACCTTGGCAATGTT 3180  
DB 3561 ACCAATATGGCACTGCCGAAATTAATATACCAAGGAGTGGTAAACCTTGGCAATGTT 3620  
QY 3181 ACCAATATGGTGAATTTAAACATTTACCCTACCGTAAACGCAACCAAGAGCATCATC 3240  
DB 3621 ACCAATATGGTGAATTTAAACATTTACCCTACCGTAAACGCAACCAAGAGCATCATC 3680  
QY 3241 GCGGAGATATATCAACAAAAGGAAGCTTTAATATATACAGACAGTAAATATGATGCT 3300  
DB 3681 GCGGAGATATATCAACAAAAGGAAGCTTTAATATATACAGACAGTAAATATGATGCT 3740  
QY 3301 GAAATCCAAATTTGGGGCAATATCTCGCAAAAAGAGGCAACCTCACGATTTCTCCCAT 3360  
DB 3741 GAAATCCAAATTTGGGGCAATATCTCGCAAAAAGAGGCAACCTCACGATTTCTCCCAT 3800  
QY 3361 AAAATTAATATCAACCAACAGATATACATCAAAAGGGTATTTGATGGAGGAGCTCTAGT 3420  
DB 3801 AAAATTAATATCAACCAACAGATATACATCAAAAGGGTATTTGATGGAGGAGCTCTAGT 3860  
QY 3421 TCAGATGCGACAAGTAAATGCCAATCTTAATTTAAACCAAGAGATTTGAAATTTGACAGAA 3480  
DB 3861 TCAGATGCGACAAGTAAATGCCAATCTTAATTTAAACCAAGAGATTTGAAATTTGACAGAA 3920  
QY 3481 GACCTAAGTATTTTCAGGTTTTCATTAAGCAGAGATTTACAGCCCAAGATGGTAGAGATTTA 3540

DB 3921 GACCTAAGTATTTTCAGGTTTTCATTAAGCAGAGATTTACAGCCAAAGATGGTAGAGATTTA 3980  
QY 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGTGGCGAAGCCAAACACACTAATCTTTAAC 3600  
DB 3981 ACTATTGGCAACAGTAATGACGGTAAACAGCGTGGCGAAGCCAAACACACTAATCTTTAAC 4040  
QY 3601 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCACAATGTGACACTAAATAGCAAGTG 3660  
DB 4041 AATGTTAAAGATTTCAAAAATCTCTGCTGACGGTCACAATGTGACACTAAATAGCAAGTG 4100  
QY 3661 AAAACATCTAGCAGCAATGGCGGAGCTGAAACCAATAGGCAACACATACCGGCTTAAC 3720  
DB 4101 AAAACATCTAGCAGCAATGGCGGAGCTGAAACCAATAGGCAACACATACCGGCTTAAC 4160  
QY 3721 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCTCAAAAAGTAAATATC 3780  
DB 4161 ATTACTGCAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCTCAAAAAGTAAATATC 4220  
QY 3781 ACCGGTGGGAAAGGTTTACCACACAGAGCTCGACCATTTAACGCAACAAATGCAAA 3840  
DB 4221 ACCGGTGGGAAAGGTTTACCACACAGAGCTCGACCATTTAACGCAACAAATGCAAA 4280  
QY 3841 GCAAGTATTTACAAACCAACAGGTGATATCAGGGTACCATTTCCGGTAAACAGGTAAGT 3900  
DB 4281 GCAAGTATTTACAAACCAACAGGTGATATCAGGGTACCATTTCCGGTAAACAGGTAAGT 4340  
QY 3901 GTTAGCGGACCTGGTGATTTTAAACCACTAAATCCGGTCAAAAATTTGAAGCGAAATCGGGT 3960  
DB 4341 GTTAGCGGACCTGGTGATTTTAAACCACTAAATCCGGTCAAAAATTTGAAGCGAAATCGGGT 4400  
QY 3961 GAGGCTAATGTAACAAAGTCAACAGGTAGATATGCGGGTACCAATTTCCGGTAAACAGGTA 4020  
DB 4401 GAGGCTAATGTAACAAAGTCAACAGGTAGATATGCGGGTACCAATTTCCGGTAAACAGGTA 4460  
QY 4021 AATGTTACGGCAAAAGCTGGCGATTTTAAACAGTTTGGGAATGGCGCAGAAATTAATGCGACA 4080  
DB 4461 AATGTTACGGCAAAAGCTGGCGATTTTAAACAGTTTGGGAATGGCGCAGAAATTAATGCGACA 4520  
QY 4081 GAAGAGCTGCAACCTTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140  
DB 4521 GAAGAGCTGCAACCTTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4580  
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTTGGGCTCAGAAATGGTTCAGATTCGCGAAGC 4200  
DB 4581 ATCACTTCAACTAAGGCTCAGGTAGACCTTTGGGCTCAGAAATGGTTCAGATTCGCGAAGC 4640  
QY 4201 ATTAATGCTGCTAATGTGACATTTAAATCTACAGGCACCTTTAACCCAGCTGCGAGCTCG 4260  
DB 4641 ATTAATGCTGCTAATGTGACATTTAAATCTACAGGCACCTTTAACCCAGCTGCGAGCTCG 4700  
QY 4261 GATATTTAAAGCAACCGCGGACCTTTGGTTTATTAACGCAAAAGATGCTTAAGCTAAATGGT 4320  
DB 4701 GATATTTAAAGCAACCGCGGACCTTTGGTTTATTAACGCAAAAGATGCTTAAGCTAAATGGT 4760  
QY 4321 GATGATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACGCAAGCGGCTCTGCTAGTGTG 4380  
DB 4761 GATGATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACGCAAGCGGCTCTGCTAGTGTG 4820  
QY 4381 ACTGGGCAACCTCAAGCAGTGTGAATATCTACTGCGGATTTAAACACAGTAAATGGGTTA 4440  
DB 4821 ACTGGGCAACCTCAAGCAGTGTGAATATCTACTGCGGATTTAAACACAGTAAATGGGTTA 4880  
QY 4441 AATATCATTTCCAAAGATGTAGAAACCTGTGCGCTTAAGAGCAAGGAAATTTAGGCTG 4500  
DB 4881 AATATCATTTCCAAAGATGTAGAAACCTGTGCGCTTAAGAGCAAGGAAATTTAGGCTG 4940  
QY 4501 AAATATATCCAGCCAGGTGTAGCAAGTGTAGAGAAAGTAAATTTGAAGCGAAACCGCTCT 4560  
DB 4941 AAATATATCCAGCCAGGTGTAGCAAGTGTAGAGAAAGTAAATTTGAAGCGAAACCGCTCT 5000  
QY 4561 GAAAAGTAAAAGATTTTCTGATGAAGAAAGAGAACATTTAGCTAACTGCTGCTAAAGT 4620



QY 1441 GGCATTCAATTAGCAAGAAAGAACTCTTTAGAAAAAGGCTCAACCAATCAATGATCAGGC 1500  
DB 1881 GGCATTCAATTAGCAAGAAAGAACTCTTTAGAAAAAGGCTCAACCAATCAATGATCAGGC 1940  
QY 1501 AAGAAAAAGGCGGCGCTATTGCTGGGCGGATATTCGGTAAATTGACGGCAATATT 1560  
DB 1941 AAGAAAAAGGCGGCGCTATTGCTGGGCGGATATTCGGTAAATTGACGGCAATATT 2000  
QY 1561 AAGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTGGTGGAGACATCGGGGCAT 1620  
DB 2001 AAGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGTGGTGGAGACATCGGGGCAT 2060  
QY 1621 TATTTATCCATTACAGCAATGCAATTTGTTAAACAAAAGAGTGGTGGTAGACCCCTGAT 1680  
DB 2061 TATTTATCCATTACAGCAATGCAATTTGTTAAACAAAAGAGTGGTGGTAGACCCCTGAT 2120  
QY 1681 GATGTAACAATTGAAGCCGAAGACCCCTTCGCAATTAATACCGGTAAATGATGAATTC 1740  
DB 2121 GATGTAACAATTGAAGCCGAAGACCCCTTCGCAATTAATACCGGTAAATGATGAATTC 2180  
QY 1741 CCACAGGACCGGTGAAGCAAGGACCCCTAAACAAAATAGCGAACTCAAAAACAGCTA 1800  
DB 2181 CCACAGGACCGGTGAAGCAAGGACCCCTAAACAAAATAGCGAACTCAAAAACAGCTA 2240  
QY 1801 ACCAATACAACTATTTCAAATTTATCTGAAAAACGCTCGACAAATGAATATACGGCATCA 1860  
DB 2241 ACCAATACAACTATTTCAAATTTATCTGAAAAACGCTCGACAAATGAATATACGGCATCA 2300  
QY 1861 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 1920  
DB 2301 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCCTCAT 2360  
QY 1921 AGTAAAGGTACGCGTGGGAGGCGCTCAGATGATGAGAGATATTACTTCAAGCGGGA 1980  
DB 2361 AGTAAAGGTACGCGTGGGAGGCGCTCAGATGATGAGAGATATTACTTCAAGCGGGA 2420  
QY 1981 AATTAAACCTTTATCTGGGAGTGGTGTGATGTTCAATAAAATATTACGCTTCATCAG 2040  
DB 2421 AATTAAACCTTTATCTGGGAGTGGTGTGATGTTCAATAAAATATTACGCTTCATCAG 2480  
QY 2041 GGTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGTGGGAATTAACAAAGCAGCG 2100  
DB 2481 GGTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGTGGGAATTAACAAAGCAGCG 2540  
QY 2101 GACGCGCAATGCTAAATTTGTCCCGCAGGCACTGTAAACCATTAACAGGAGGGA 2160  
DB 2541 GACGCGCAATGCTAAATTTGTCCCGCAGGCACTGTAAACCATTAACAGGAGGGA 2600  
QY 2161 GATTTCAGGGCTAACAGGTATCTTTTAAACGGAACGGGTAAAGTCTGAATATCATTTCA 2220  
DB 2601 GATTTCAGGGCTAACAGGTATCTTTTAAACGGAACGGGTAAAGTCTGAATATCATTTCA 2660  
QY 2221 TCAGTGAATATTAAACCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280  
DB 2661 TCAGTGAATATTAAACCAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2720  
QY 2281 ATTAACCAAACTACGAGAAAGAACCTCTGTTTGGCAACCGCCATGATTCGCACTGG 2340  
DB 2721 ATTAACCAAACTACGAGAAAGAACCTCTGTTTGGCAACCGCCATGATTCGCACTGG 2780  
QY 2341 AACGTGAGTCTCTTATCTAGAGACAGGCGCAATTTTACCTTTTAAATACATTTCA 2400  
DB 2781 AACGTGAGTCTCTTATCTAGAGACAGGCGCAATTTTACCTTTTAAATACATTTCA 2840  
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAGCTCTCGAGGGTGAATTTTAAACGGC 2460  
DB 2841 AGCAATAGCAAGGCTTAAACACACAGTATAGAGCTCTCGAGGGTGAATTTTAAACGGC 2900  
QY 2461 GTAATGGCAACATGTCATTTCAATCTCAAGAGAGGCGAAAGTAAATTTCAATTTAAA 2520  
DB 2901 GTAATGGCAACATGTCATTTCAATCTCAAGAGAGGCGAAAGTAAATTTCAATTTAAA 2960

QY 2521 CCAACGAGAACATGAACACAGCAAAACCTTTACCAATTCGGTTCGTTAGCCAAATATCACA 2580  
DB 2961 CCAACGAGAACATGAACACAGCAAAACCTTTACCAATTCGGTTCGTTAGCCAAATATCACA 3020  
QY 2581 GCCACTGGTGGGCGCTCTGTTTTTTGATATATATGCCAACCAATCTGCGAGAGGGCT 2640  
DB 3021 GCCACTGGTGGGCGCTCTGTTTTTTGATATATATGCCAACCAATCTGCGAGAGGGCT 3080  
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTTAAATTCOCAT 2700  
DB 3081 GAGTTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTTAAATTCOCAT 3140  
QY 2701 GTTCGGCGGATCAGCCTTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATTTCA 2760  
DB 3141 GTTCGGCGGATCAGCCTTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATTTCA 3200  
QY 2761 AATTTCAAGCTCAGACAGCAAGAAAGTATTTTATGACGGGTACGCACCAATGCCATC 2820  
DB 3201 AATTTCAAGCTCAGACAGCAAGAAAGTATTTTATGACGGGTACGCACCAATGCCATC 3260  
QY 2821 AATTTCAAGCTCAGACCAATATCCATCTCGGGGTAAATGTCACCCCTGGTGGACAAAATCA 2880  
DB 3261 AATTTCAAGCTCAGACCAATATCCATCTCGGGGTAAATGTCACCCCTGGTGGACAAAATCA 3320  
QY 2881 AGCAGCAGCATTACGGGGAATATTACTATCGAGAAAGCAGCAATGTTACGCTAGAGCC 2940  
DB 3321 AGCAGCAGCATTACGGGGAATATTACTATCGAGAAAGCAGCAATGTTACGCTAGAGCC 3380  
QY 2941 AATAAGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGCCT 3000  
DB 3381 AATAAGCCCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGCCT 3440  
QY 3001 GTTAATGGAGTTAAGTTTAACTGGCGAAATGCAAGATATAAGGCAATCTCACTATT 3060  
DB 3441 GTTAATGGAGTTAAGTTTAACTGGCGAAATGCAAGATATAAGGCAATCTCACTATT 3500  
QY 3061 TCAGAAAGCGCCACTTTTAAAGAAAGACTAGAGATACCTTAATATACCGCAATTTT 3120  
DB 3501 TCAGAAAGCGCCACTTTTAAAGAAAGACTAGAGATACCTTAATATACCGCAATTTT 3560  
QY 3121 ACCAATATGGCAGTCCGAAATTAATATACACAGGAGTGTAAACTTGGCAATGTT 3180  
DB 3561 ACCAATATGGCAGTCCGAAATTAATATACACAGGAGTGTAAACTTGGCAATGTT 3620  
QY 3181 ACCAATATGTTGATTTAAACATTAACCACTACCGTAAACGCAACCAAGAGCATCATC 3240  
DB 3621 ACCAATATGTTGATTTAAACATTAACCACTACCGTAAACGCAACCAAGAGCATCATC 3680  
QY 3241 GCGGAGATATAATCAACAAAGAAAGAGCTTAATATTTACAGACAGTAATGATGCT 3300  
DB 3681 GCGGAGATATAATCAACAAAGAAAGAGCTTAATATTTACAGACAGTAATGATGCT 3740  
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAAAGCAACCTCACGATTTCTCCGAT 3360  
DB 3741 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAAAGCAACCTCACGATTTCTCCGAT 3800  
QY 3361 AAAATTAATATCAACAAAGAGATTAACATCAAAAGGTTATGAGAGAGCTTCTAGT 3420  
DB 3801 AAAATTAATATCAACAAAGAGATTAACATCAAAAGGTTATGAGAGAGCTTCTAGT 3860  
QY 3421 TCAGATGCGCAAGTAAATGCCAAGCTTAACATTAACCAAGAAATTCGAATTTGACAGAA 3480  
DB 3861 TCAGATGCGCAAGTAAATGCCAAGCTTAACATTAACCAAGAAATTCGAATTTGACAGAA 3920  
QY 3481 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGATTAACAGCCAAAGATGTTAGAGATTTA 3540  
DB 3921 GACCTAAGTATTTTCAGGTTTCAATTAAGACAGATTAACAGCCAAAGATGTTAGAGATTTA 3980  
QY 3541 ACTATTGGCAACAGTATGAGCGGTAAACAGCGGTGCCGAAAGCAACAGTAACTTTTAAAC 3600  
DB 3981 ACTATTGGCAACAGTATGAGCGGTAAACAGCGGTGCCGAAAGCAACAGTAACTTTTAAAC 4040  
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCACCAATGTGACACTTAATAGCAAGTG 3660

us-10-092-880-3.rge

Tue Mar 25 08:39:00 2003

Db	4041	AAATGTTAAAGATTCACAAATCTCTGCTGAGGCTCACAATGTGACACTAAATATGCAAGATG	4100	5121	ACCAGACCGTCAAGTCAAGTGATAATTTCTGAAGGTAGGCGTGTCTTCAAGTGTGAAT	5180
Qy	3661	AAAAATCTAGAGCAATGGCGAGCTGAAGCAATACGACAAAGATACCGGCTTAAGT	3720	4741	GGCCACAGATATGTACCAATGTTGCTGACGATGGACAGCCGTAGTCAATATGACAAAG	4800
Db	4101	AAAAATCTAGAGCAATGGCGAGCTGAAGCAATACGACAAAGATACCGGCTTAAGT	4160	5181	GGCCACAGATATGTACCAATGTTGCTGACGATGGACAGCCGTAGTCAATATGACAAAG	5240
Qy	3721	ATTACTGCAAAATGATGAGTAAGTAACAAGATATTACTTCTCTCAAAACAGTAAATATC	3780	4801	GTAGATTTTCACTCCCGCAATGAGTCATTTATTTTCGATATTTACTGTGGGTAAA	4860
Db	4161	ATTACTGCAAAATGATGAGTAAGTAACAAGATATTACTTCTCTCAAAACAGTAAATATC	4220	5241	GTAGATTTTCACTCCCGCAATGAGTCATTTATTTTCGATATTTACTGTGGGTAAA	5300
Qy	3781	ACCGGCTCGGAAAGGTTACACACAGAGCTCGACCATTAACGCAACAATGGCAAA	3840	4861	GTTCAAGTACGGGCTTTTACCACATCTGTAAATAATTCGAGAAATACATAAAGATATTTT	4920
Db	4221	ACCGGCTCGGAAAGGTTACACACAGAGCTCGACCATTAACGCAACAATGGCAAA	4280	5301	GTTCAAGTACGGGCTTTTACCACATCTGTAAATAATTCGAGAAATACATAAAGATATTTT	5360
Qy	3841	GCAAGTATTACAAACAAACAGGTTATACAGGGTACGATTTCCGGTAAACCGGTAAGT	3900	4921	AACAGGTTATTATTATG	4937
Db	4281	GCAAGTATTACAAACAAACAGGTTATACAGGGTACGATTTCCGGTAAACCGGTAAGT	4340	5361	AACAGGTTATTATTATG	5377
Qy	3901	GTTAGCGGACTGGTGAATTAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATCGGT	3960	RESULT 12		
Db	4341	GTTAGCGGACTGGTGAATTAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATCGGT	4400	LOCUS		
Qy	3961	GAGGCTAATGTAAACAGTGAACAGGTAACAATTTGGCGGTACAAATTTCCGGTAAACGGTA	4020	DEFINITION		
Db	4401	GAGGCTAATGTAAACAGTGAACAGGTAACAATTTGGCGGTACAAATTTCCGGTAAACGGTA	4460	SEQUENCE		
Qy	4021	AAATGTTAAAGATTCACAAATCTCTGCTGAGGCTCACAATGTGACACTAAATATGCAAGATG	4080	ACCESSION		
Db	4461	AAATGTTAAAGATTCACAAATCTCTGCTGAGGCTCACAATGTGACACTAAATATGCAAGATG	4520	VERSION		
Qy	4081	GAAGGCTGCAACCTTAAACCGCAACAGGTAACAATTTGGCGGTACAAATTTCCGGTAAACGGTA	4580	KEYWORDS		
Db	4521	GAAGGCTGCAACCTTAAACCGCAACAGGTAACAATTTGGCGGTACAAATTTCCGGTAAACGGTA	4640	SOURCE		
Qy	4141	ATCACTTCACTAAGGTCAGTACAGTACCTTGGCTCAGAAATGAGTATGCAAGGACG	4200	ORGANISM		
Db	4581	ATCACTTCACTAAGGTCAGTACAGTACCTTGGCTCAGAAATGAGTATGCAAGGACG	4640	REFERENCE		
Qy	4201	ATTAATGCTGTAATGTCACATTAATTAACACTACAGGCACTTAACACCGTGGCAGGCTG	4260	AUTHORS		
Db	4641	ATTAATGCTGTAATGTCACATTAATTAACACTACAGGCACTTAACACCGTGGCAGGCTG	4700	TITLE		
Qy	4261	GATATTAAAGCAACCGGCACTTGGTTATTAAACGCAAAAGATGTAAGCTAAATGGT	4320	JOURNAL		
Db	4701	GATATTAAAGCAACCGGCACTTGGTTATTAAACGCAAAAGATGTAAGCTAAATGGT	4760	FEATURES		
Qy	4321	GATGATCAGGTCAGTACAGTACAGTACCTTGGCTCAGAAATGAGTATGCAAGGACG	4380	BASE COUNT		
Db	4761	GATGATCAGGTCAGTACAGTACAGTACCTTGGCTCAGAAATGAGTATGCAAGGACG	4820	ORIGIN		
Qy	4381	ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA	4440	Query Match		
Db	4821	ACTGGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA	4880	Best Local Similarity		
Qy	4441	AAATATATCCAGCCAGGTCAGTACAGTACCTTGGCTTAAAGGCAAGCAATTAAGGTC	4500	Matches		
Db	4881	AAATATATCCAGCCAGGTCAGTACAGTACCTTGGCTTAAAGGCAAGCAATTAAGGTC	4940	4864; Conservative		
Qy	4501	AAATATATCCAGCCAGGTCAGTACAGTACCTTGGCTTAAAGGCAAGCAATTAAGGTC	4560	Qy 1		
Db	4941	AAATATATCCAGCCAGGTCAGTACAGTACCTTGGCTTAAAGGCAAGCAATTAAGGTC	5000	Db 441		
Qy	4561	AAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTTTGGTGAAGT	4620	Qy 61		
Db	5001	AAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTTTGGTGAAGT	5060	Db 441		
Qy	4621	GCTGTAGTGTGTTGAGGCAAAATTAACAATACAGTCAATACACAAATGAATTTACA	4680	Qy 561		
Db	5061	GCTGTAGTGTGTTGAGGCAAAATTAACAATACAGTCAATACACAAATGAATTTACA	5120	Db 501		
Qy	4681	ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAAGGCGTGTCTCAAGTGGTAAT	4740	Qy 121		
Db	5120	ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAAGGCGTGTCTCAAGTGGTAAT	5180	Db 561		
Qy	4741	AAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTTTGGTGAAGT	4800	Qy 181		
Db	5180	AAAAAGTAAAAAGATTTATCTGATGAAGAAAGAAACATTAAGCTTTGGTGAAGT	5240	Db 621		
Qy	4801	ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTGCTGCTGAATTTGCA	4860	Qy 561		
Db	5240	ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTGCTGCTGAATTTGCA	5300	Qy 181		
Qy	4861	CGGGTGTGACCAATTCACAGAAAGGCTTCGGCTATGTTACTATCTTTAGGTGAAC	4920	Db 621		
Db	5300	CGGGTGTGACCAATTCACAGAAAGGCTTCGGCTATGTTACTATCTTTAGGTGAAC	5360	Qy 621		
Qy	4921	AACAGGTTATTATTATG	4937	Qy 181		
Db	5361	AACAGGTTATTATTATG	5377	Db 621		



Db 861 CGGGTGTGTGACCAATCCACAGAAAAAGGAGCGAAAAACCTGCTCGCATGAAAGTGCGT 920  
Qy 481 CACTTAGGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540  
Db 921 CACTTAGGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 980  
Qy 541 CAATCTGTTTACCAAGCGGCTTACAGGAATGGATGTAGTACACGGGCACAGCCACTATG 600  
Db 981 CAATCTGTTTACCAAGCGGCTTAAATGATGATGATGATGATGATGATGATGATGATGAT 1040  
Qy 601 CAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 660  
Db 1041 AGAAAAACAAAGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 1100  
Qy 661 CAATTTAATCATCGACCAAAATGAATGGTGGAGTTTAAAGAAAAACAACACTCGGC 720  
Db 1101 CAATTTAATCATCGACCAAAATGAATGGTGGAGTTTAAAGAAAAACAACACTCGGC 1160  
Qy 721 GTATTTCAACCGTGTACATCAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC 780  
Db 1161 GTATTTCAACCGTGTACATCAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC 1220  
Qy 781 GGCAAGTCTTTTAAATCAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC 840  
Db 1221 GGCAAGTCTTTTAAATCAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTAAC 1280  
Qy 841 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTTAACGAAACATCAAGCGCGTAAT 900  
Db 1281 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTTAACGAAACATCAAGCGCGTAAT 940  
Qy 901 TACACCTTCGAGCAACCAACCAATCTTAACGAAACATCAAGCGCGTAAT 1340  
Db 1341 TACACCTTCGAGCAACCAACCAATCTTAACGAAACATCAAGCGCGTAAT 1400  
Qy 961 ACTGTCGTTAAAGCGGAGTGAATCTTAAATGTTGGCAAGTGAATGGAATCAACGCGT 1020  
Db 1401 ACTGTCGTTAAAGCGGAGTGAATCTTAAATGTTGGCAAGTGAATGGAATCAACGCGT 1060  
Qy 1021 ATTACGTTAAATGTTGGAGCATTTCTTAAATGTTGGCAAGTGAATGGAATCAACGCGT 1080  
Db 1461 ATTACGTTAAATGTTGGAGCATTTCTTAAATGTTGGCAAGTGAATGGAATCAACGCGT 1140  
Qy 1081 ATAATAACCAACCAATTTACTTACAGCATTTGCGCGCTGAAATGAAGCGTCAATCTG 1180  
Db 1521 ATAATAACCAACCAATTTACTTACAGCATTTGCGCGCTGAAATGAAGCGTCAATCTG 1220  
Qy 1141 GCGCATATTTTGCAGAGCGGTAACATTAATGTCGCTGCTGCAATTAAGCGTCAATCTG 1260  
Db 1581 GCGCATATTTTGCAGAGCGGTAACATTAATGTCGCTGCTGCAATTAAGCGTCAATCTG 1300  
Qy 1201 GGTAAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCGGCAATTAATGTTCTTCCGCC 1340  
Db 1641 GGTAAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCGGCAATTAATGTTCTTCCGCC 1380  
Qy 1261 AAAGAGGTGAAGCGGCAATTTGCGGTGTAATTTCCGCTCAAAATCAACGCTAAAGGC 1420  
Db 1701 AAAGAGGTGAAGCGGCAATTTGCGGTGTAATTTCCGCTCAAAATCAACGCTAAAGGC 1460  
Qy 1321 GCGAAGCTGATGATTACAGCGGTAATGTCACATTAATAACAGGTGCAATTTATCGACCT 1500  
Db 1761 GCGAAGCTGATGATAAGTCCGATTAAGTCAATTAATAACAGGTGCAATTTATCGACCT 1540  
Qy 1381 TCAGTAAGAGGGGAGAACTTACCTTGGGGTGACGAGCGCGGCAAGGTAAAC 1600  
Db 1821 TCAGTAAGAGGGGAGAACTTACCTTGGGGTGACGAGCGCGGCAAGGTAAAC 1640  
Qy 1441 GGCATTCATTAAGCAAAAGAACTTCTTTAGAAAAGGTCAACCATCAATGATCAGGC 1680  
Db 1881 GGCATTCATTAAGCAAAAGAACTTCTTTAGAAAAGGTCAACCATCAATGATCAGGC 1720  
Qy 1501 AAAGAAAAAGCGGCGCTATTGTTGGGCGATATTGCGGTAAATGACGCGCAATATT 1560  
Db 1941 AAAGAAAAAGCGGCGCTATTGTTGGGCGATATTGCGGTAAATGACGCGCAATATT 2000

Qy 1561 AACGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTGTGTTGAGACATCGGGGAT 1620  
Db 2001 AACGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTGTGTTGAGACATCGGGGAT 2060  
Qy 1621 TATTTATCCATTGACAGCAATGCAATGTTTAAACAAAGAGTGGTGTGTTAGACCCCTGAT 1680  
Db 2061 TATTTATCCATTGACAGCAATGCAATGTTTAAACAAAGAGTGGTGTGTTAGACCCCTGAT 2120  
Qy 1681 GATGTAACAAATTTGAAGCGCAAGACCCCTTCGCAATTAATACCGGTATAAATGATGATTC 1740  
Db 2121 GATGTAACAAATTTGAAGCGCAAGACCCCTTCGCAATTAATACCGGTATAAATGATGATTC 2180  
Qy 1741 CCAACAGCGCCGTTGAAGCAAGCGCCCTTAAAAAAAATAGCAACTCAAAACAGGCTA 1800  
Db 2181 CCAACAGCGCCGTTGAAGCAAGCGCCCTTAAAAAAAATAGCAACTCAAAACAGGCTA 2240  
Qy 1801 ACCAATCAACTATTTCAAAATTTCTGAAACCGCTGGACAAATGAATATAACGGCATCA 1860  
Db 2241 ACCAATCAACTATTTCAAAATTTCTGAAACCGCTGGACAAATGAATATAACGGCATCA 2300  
Qy 1861 AGAAACCTTACCCTTAAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 1920  
Db 2301 AGAAACCTTACCCTTAAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTCCTCAT 2360  
Qy 1921 AGTAAAGGTGAGCGTGGCGGAGCGCTTCAAGATTGATGAGATTAATTTCTTAAAGCGGA 1980  
Db 2361 AGTAAAGGTGAGCGTGGCGGAGCGCTTCAAGATTGATGAGATTAATTTCTTAAAGCGGA 2420  
Qy 1981 AATTTAAACCAATTTATTTACCGCGCTTCCGCTAGCTTTTGAAGTGGAAATAACAAAGCACGC 2480  
Db 2421 AATTTAAACCAATTTATTTACCGCGCTTCCGCTAGCTTTTGAAGTGGAAATAACAAAGCACGC 2540  
Qy 2041 GGTTTTAAATATTTACCGCGCTTCCGCTAGCTTTTGAAGTGGAAATAACAAAGCACGC 2600  
Db 2481 GGTTTTAAATATTTACCGCGCTTCCGCTAGCTTTTGAAGTGGAAATAACAAAGCACGC 2660  
Qy 2101 GACGCGCAATGCTAAATTTGCGCCAGCGGCACTGTAACCAATTACAGGAGAGGAA 2160  
Db 2541 GACGCGCAATGCTAAATTTGCGCCAGCGGCACTGTAACCAATTACAGGAGAGGAA 2600  
Qy 2161 GATTTACAGGCTTAAACCAATTTTAAACGGAACGGGTAAAGGCTGAATATCATTTCA 2220  
Db 2601 GATTTACAGGCTTAAACCAATTTTAAACGGAACGGGTAAAGGCTGAATATCATTTCA 2660  
Qy 2221 TCAGTCAATTAATTTAAACCAATTTTAAACGGAACGGGTAAAGGCTGAATATCATTTCA 2280  
Db 2661 TCAGTCAATTAATTTAAACCAATTTTAAACGGAACGGGTAAAGGCTGAATATCATTTCA 2720  
Qy 2281 ATTAACCAACTACGAGAAAGCAACCTCGTATTTGGCAACCAAGCTGATTCGCACTGG 2340  
Db 2721 ATTAACCAACTACGAGAAAGCAACCTCGTATTTGGCAACCAAGCTGATTCGCACTGG 2780  
Qy 2341 AACGTCAGTCTTAAATCTAGACAGCGGCAATTTTACCTTTTAAATACATTTCA 2400  
Db 2781 AACGTCAGTCTTAAATCTAGACAGCGGCAATTTTACCTTTTAAATACATTTCA 2840  
Qy 2401 AGCAATAGCAAGGCTTAAACCAATTTTAAACGGAACGGGTAAAGGCTGAATATCATTTCA 2460  
Db 2841 AGCAATAGCAAGGCTTAAACCAATTTTAAACGGAACGGGTAAAGGCTGAATATCATTTCA 2900  
Qy 2461 GTAAATGGCAACTGTCATTTCAATCTCAAGAGAGGAGGCAAGTTAATTTCAAAATTA 2520  
Db 2901 GTAAATGGCAACTGTCATTTCAATCTCAAGAGAGGAGGCAAGTTAATTTCAAAATTA 2960  
Qy 2521 CCAACAGAGCAATGACACCAAGCAACCTTTACCAATTTGCGGTGTTTACCCATATCACA 2580  
Db 2961 CCAACAGAGCAATGACACCAAGCAACCTTTACCAATTTGCGGTGTTTACCCATATCACA 3020  
Qy 2581 GGCACGTGGGGGCTCTGTTTTTTTGTATATATATGCAACCAATTTGCGGAGAGGCT 2640  
Db 3021 GGCACGTGGGGGCTCTGTTTTTTTGTATATATATGCAACCAATTTGCGGAGAGGCT 3080

us-10-092-880-3.rge

Tue Mar 25 08:39:00 2003

QY	2641	GAGTTAAATGAGTGAATTAATCTCTAAGGGGCTAAATTTACCTTAATTTCCCAT	2700
Db	3081	GAGTTAAATGAGTGAATTAATCTCTAAGGGGCTAAATTTACCTTAATTTCCCAT	3140
QY	2701	GTTCGGCGGATGAGCTTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATCA	2760
Db	3141	GTTCGGCGGATGAGCTTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATCA	3200
QY	2761	AATTTGAGCTCAGACAGCAAGAGATGATTTTATGACGGGTACGCAACCAATGCGATC	2820
Db	3201	AATTTGAGCTCAGACAGCAAGAGATGATTTTATGACGGGTACGCAACCAATGCGATC	3260
QY	2821	AATTCACCTTACACATATCCATCTGGCGGTAAATGTACCCCTTGGTGGACAAACTCA	2880
Db	3261	AATTCACCTTACACATATCCATCTGGCGGTAAATGTACCCCTTGGTGGACAAACTCA	3320
QY	2881	AGCAGCAGCATACGGGGATATCTATCAGAAAGAGAGCAAAATGTACGCTAGAAGCC	2940
Db	3321	AGCAGCAGCATACGGGGATATCTATCAGAAAGAGAGCAAAATGTACGCTAGAAGCC	3380
QY	2941	AATAACGCCCTTAATCAGCAAAACATAGGGATAGAGTTATATAAACTTTGGCAGCTTGCTC	3000
Db	3381	AATAACGCCCTTAATCAGCAAAACATAGGGATAGAGTTATATAAACTTTGGCAGCTTGCTC	3440
QY	3001	GTATATGGGAGTTTAACTGGCGAAATGCGAATATGAGATATTAAGGCAATCTCACTATT	3060
Db	3441	GTATATGGGAGTTTAACTGGCGAAATGCGAATATGAGATATTAAGGCAATCTCACTATT	3500
QY	3061	TCAGAAAGCGGCACCTTTTAAAGGAAAGCTAGAGATACCTTAATATACCGGCAATTTT	3120
Db	3501	TCAGAAAGCGGCACCTTTTAAAGGAAAGCTAGAGATACCTTAATATACCGGCAATTTT	3560
QY	3121	ACCAATATGGCAGCTGCGGAAATTAATATACACAGAGGTGGTAAACTTGGCAATGTT	3180
Db	3561	ACCAATATGGCAGCTGCGGAAATTAATATACACAGAGGTGGTAAACTTGGCAATGTT	3620
QY	3181	ACCAATATGGTGTATTAACATTAACACTACCGCTTAACGCAACCAAGCAATCATC	3240
Db	3621	ACCAATATGGTGTATTAACATTAACACTACCGCTTAACGCAACCAAGCAATCATC	3680
QY	3241	GGCGGAGATATATCAACAAAAAGGAAAGCTTAAATATTAACAGACATATATGATGCT	3300
Db	3681	GGCGGAGATATATCAACAAAAAGGAAAGCTTAAATATTAACAGACATATATGATGCT	3740
QY	3301	GAAATCCAAATGGCGCAATATCTCGCAAAAGGAAAGCAACCTCACGATTTCTTCGAT	3360
Db	3741	GAAATCCAAATGGCGCAATATCTCGCAAAAGGAAAGCAACCTCACGATTTCTTCGAT	3800
QY	3361	AAATTAATATCAACCAACAGATAACCAATCAAAAGGGTATTTGATGGAGAGACTTAGT	3420
Db	3801	AAATTAATATCAACCAACAGATAACCAATCAAAAGGGTATTTGATGGAGAGACTTAGT	3860
QY	3421	TCAGATGGCAAGTATGCAACCTTAACCTTAAATCAAAAGGAAAGTGAATTTGACAGAA	3480
Db	3861	TCAGATGGCAAGTATGCAACCTTAACCTTAAATCAAAAGGAAAGTGAATTTGACAGAA	3920
QY	3481	GACCTAAGTATTTGAGTTTCAATTAAGAGAGATTAACGCCAAAGAGTGGTAGAGATTA	3540
Db	3921	GACCTAAGTATTTGAGTTTCAATTAAGAGAGATTAACGCCAAAGAGTGGTAGAGATTA	3980
QY	3541	ACTATTGGCAACAGTAAATGAGGTACAGCGGTCCGCAAGGCAACCACTTAATTTTAAAC	3600
Db	3981	ACTATTGGCAACAGTAAATGAGGTACAGCGGTCCGCAAGGCAACCACTTAATTTTAAAC	4040
QY	3601	AATGTTAAAGATTCAAAATCTCTGCTACCGTCAATGTGACACTTAATATAGCAAGTG	3660
Db	4041	AATGTTAAAGATTCAAAATCTCTGCTACCGTCAATGTGACACTTAATATAGCAAGTG	4100
QY	3661	AAAACATCTAGCAGCAATGGCGGAGTGAAGCAATAGCGACAGCATACCGGCTTAAC	3720
Db	4101	AAAACATCTAGCAGCAATGGCGGAGTGAAGCAATAGCGACAGCATACCGGCTTAAC	4160
QY	3721	ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCAACAAAGTAAATATC	3780
Db	4161	ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTTACTTCTCAACAAAGTAAATATC	4220
QY	3781	ACCGCGTCGAAAGAGTTTACACACAGCAGGCTGACCATTAACGCAACAAATGGCAAA	3840
Db	4221	ACCGCGTCGAAAGAGTTTACACACAGCAGGCTGACCATTAACGCAACAAATGGCAAA	4280
QY	3841	GCAAGTATTACAAACCAACAGGTGATATACGCGTACGATTTCCGCTTAACACCGTAAGT	3900
Db	4281	GCAAGTATTACAAACCAACAGGTGATATACGCGTACGATTTCCGCTTAACACCGTAAGT	4340
QY	3901	GTTAGCGGACTGGTGAATTTAACCACTAAATCGGCTCAAAATTTGAAGCAAAATCGGCT	3960
Db	4341	GTTAGCGGACTGGTGAATTTAACCACTAAATCGGCTCAAAATTTGAAGCAAAATCGGCT	4400
QY	3961	GAGGCTAATGTAAACAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGCTTAACGTA	4020
Db	4401	GAGGCTAATGTAAACAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGCTTAACGTA	4460
QY	4021	ATGTTTACGCGCAACCGCTTGGCGATTTAACAGTTGGGAATGGCCGCAAAATTTATCGACA	4080
Db	4461	ATGTTTACGCGCAACCGCTTGGCGATTTAACAGTTGGGAATGGCCGCAAAATTTATCGACA	4520
QY	4081	GAAGGAGTGAACCTTAAACCGCAACAGGGAATACCTTGAATGAGCCCGGTTCTAGC	4140
Db	4521	GAAGGAGTGAACCTTAAACCGCAACAGGGAATACCTTGAATGAGCCCGGTTCTAGC	4580
QY	4141	ATCAGCTCAACTTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTATCGACATCGCAGAACG	4200
Db	4581	ATCAGCTCAACTTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTATCGACATCGCAGAACG	4640
QY	4201	ATTAAATGCTGCTTAATGTGACATTAATACTACAGGACCTTTAAACACCGTGGCAGGCTCG	4260
Db	4641	ATTAAATGCTGCTTAATGTGACATTAATACTACAGGACCTTTAAACACCGTGGCAGGCTCG	4700
QY	4261	GATATTAAGCAACACCGGCACTTTGGTTTAAACGCAAAAGATGCTAAAGCTAAATGGT	4320
Db	4701	GATATTAAGCAACACCGGCACTTTGGTTTAAACGCAAAAGATGCTAAAGCTAAATGGT	4760
QY	4321	GATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAACGCAAGCGGCTCTGGTAGTGTG	4380
Db	4761	GATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAACGCAAGCGGCTCTGGTAGTGTG	4820
QY	4381	ACTCGGCAACCTCAACGACGTGATATCAGCTGGGATTTAAACACAGTAAATGGGTTA	4440
Db	4821	ACTCGGCAACCTCAACGACGTGATATCAGCTGGGATTTAAACACAGTAAATGGGTTA	4880
QY	4441	AATATCATTTTCCGAAGATGGTAGAACAACACTGCGCTTAAAGAGGCAAGGAAATGGAGTG	4500
Db	4881	AATATCATTTTCCGAAGATGGTAGAACAACACTGCGCTTAAAGAGGCAAGGAAATGGAGTG	4940
QY	4501	AAATATATCCAGCCAGGTGATAGCAAGTGTAGAGAGTGAATTTGAAGCAAAACGCTCTT	4560
Db	4941	AAATATATCCAGCCAGGTGATAGCAAGTGTAGAGAGTGAATTTGAAGCAAAACGCTCTT	5000
QY	4561	GAAAGAGTAAAGATTTTCTGATGAAGAAAGAGAACACTTAGCTTAACCTTGGTGAAGT	4620
Db	5001	GAAAGAGTAAAGATTTTCTGATGAAGAAAGAGAACACTTAGCTTAACCTTGGTGAAGT	5060
QY	4621	GCTGTACGTTTGTGAGCCAAATTAATCAATTTACAGTCAATACACAAATGAATTTACA	4680
Db	5061	GCTGTACGTTTGTGAGCCAAATTAATCAATTTACAGTCAATACACAAATGAATTTACA	5120
QY	4681	ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAGGGCTGTTCTCAAGTGGTAAAT	4740
Db	5121	ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAGGGCTGTTCTCAAGTGGTAAAT	5180
QY	4741	GGCCACAGAGTATGTACCAATTTGCTGACGATGGACGCCGTAGTCAATTTGACAAAG	4800
Db	5181	GGCCACAGAGTATGTACCAATTTGCTGACGATGGACGCCGTAGTCAATTTGACAAAG	5240
QY	4801	GTAGATTTCACTCCATGAAGTCAATTTTATTTCTGCTATTTTATCTGCTGTTGGTGTAAA	4860
Db	5241	GTAGATTTCACTCCATGAAGTCAATTTTATTTCTGCTATTTTATCTGCTGTTGGTGTAAA	4900

```

Db 5241 GTAGATTTTCATCCTGCAATGAAGTCATTTTATTTTCGTATTATTTACTGTGGTTAAA 5300
QY 4861 GTTCAGTACGGCTTTACCCATCTTGTAAAAAATTCAGGAGATACAAATAAAGTATTTTT 4920
Db 5301 GTTCAGTACGGCTTTACCCATCTTGTAAAAAATTCAGGAGATACAAATAAAGTATTTTT 5360
QY 4921 AACAGGTATATTATG 4937
Db 5361 AACAGGTATATTATG 5377

RESULT 13
LOCUS BD010358 9323 bp DNA linear PAT 31-JAN-2002
DEFINITION High molecular weight surface protein of non-typeable haemophils.
ACCESSION BD010358
VERSION BD010358.1 GI:18638731
KEYWORDS JP 2001503602-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 9323)
AUTHORS Barenkamp, S. J.
TITLE High molecular weight surface protein of non-typeable haemophils
JOURNAL Patent: JP 2001503602-A 4 21-MAR-2001;
COMMENT SENTLOUIS UNIV, WASHINGTON UNIV
OS Haemophils
PN JP 2001503602-A/4
PD 21-MAR-2001
PF 01-APR-1997 JP 1997535346
PR 01-APR-1996 US 08/617697
PI STEPHEN J BARENKAMP
PC C07H21/02, C07H21/04, C12P21/06, A61K39/102
FH Key
FT Location/Qualifiers
FT 1. 9323
FT /organism="Haemophils",
FEATURES
source 1. 9323
BASE COUNT 3134 a 1828 c 1833 g 2528 t
ORIGIN
Query Match 97.6%; Score 4817; DB 6; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4862; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 1 TAAATATACAGATATAAAATTAATCAAGATTTTGTGATGACAAACAACAAATTACAA 60
Db 441 TAAATATACAGATATAAAATTAATCAAGATTTTGTGATGACAAACAACAAATTACAA 500
QY 61 CACCTTTTTCAGTCTATATGCAATATTTTAAAAAATAGTATAAATCCGCCATATAA 120
Db 501 CACCTTTTTCAGTCTATATGCAATATTTTAAAAAATAGTATAAATCCGCCATATAA 560
QY 121 AATGGTATAATCTTTCATCTTTTCACTTTTATCTTTTCATCTTTTCATCTTTTCAT 180
Db 561 AATGGTATAATCTTTCATCTTTTCACTTTTATCTTTTCATCTTTTCATCTTTTCAT 620
QY 181 CTTTCATCTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTTT 240
Db 621 CTTTCATCTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTTT 680
QY 241 CACATGAATGATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 681 CACATGAATGATCAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740
QY 301 GAACGCAAAATGATAAGTAAATTAATTTGTTCAACTAACCTTAGGAGAAATATGAACAAG 360
Db 741 GAACGCAAAATGATAAGTAAATTAATTTGTTCAACTAACCTTAGGAGAAATATGAACAAG 800
QY 361 ATATATCGTCTCAATTCAGCAAAACCCCTGAATGCTTTGGTGTCTGTCTGAATTGCA 420

```

```

Db 801 ATATATCTGCTCAAAATTCAGCAAAACGCTGAATCCTTTGGTGTCTGAATTTGGCA 860
QY 421 CGGGTGTGTGACCATTCACAGAAAAAGGCTTCGGCTATCTTACTATCTTTAGGTGTAAAC 480
Db 861 CGGGTGTGTGACCATTCACAGAAAAAGGCGGAAAAACCTGCTCGCATGAAGTGGCT 920
QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATCTTACTATCTTTAGGTGTAAAC 540
Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATCTTACTATCTTTAGGTGTAAAC 980
QY 541 CAATCTCTTTTAGCAAGCGCTTACAGGAATGATGTAGTACAGCGCACAGCCACTATG 600
Db 981 CAATCTCTTTTAGCAAGCGCTTACAGGAATGATGTAGTACAGCGCACAGCCACTATG 1040
QY 601 CAAGTAGATGGTAAATAAACCACTTTATCCGCAACAGTGTTCGCAAAATGAATGGTGCACTTTTAC 1100
Db 1041 CAATCTCTTTTAGCAAGCGCTTACAGGAATGATGTAGTACAGCGCACAGCCACTATG 1160
QY 661 CAATTTTACATCGACCAAAATGAATGGTGCAGTGTTCGCAAAATGAATGGTGCACTTTTAC 1220
Db 1101 CAATTTTACATCGACCAAAATGAATGGTGCAGTGTTCGCAAAATGAATGGTGCACTTTTAC 720
QY 721 GTATTCAACCGTGTGTACATCTAACCAATCTCCCAATTTAAAGAGGATTTAGATTTTAAAC 780
Db 1161 GTATTCAACCGTGTGTACATCTAACCAATCTCCCAATTTAAAGAGGATTTAGATTTTAAAC 1280
QY 781 GGACAAGTCTTTTAAATCAACCCCAATGGTATACATAGGTAAAGAGGATTTAGATTTTAAAC 840
Db 1221 GGACAAGTCTTTTAAATCAACCCCAATGGTATACATAGGTAAAGAGGATTTAGATTTTAAAC 900
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTTCTAACGAAAAACATCAAGGCGCGTAAT 960
Db 1281 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTTCTAACGAAAAACATCAAGGCGCGTAAT 1340
QY 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTGAATCACCGTTTAAAT 960
Db 1341 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTGAATCACCGTTTAAAT 1400
QY 961 ACTCTCGGTAAAGACGGCGAGTGTAAATCTTATTTGGTGGCAAAAGTGAAGAACGAGGTGTG 1020
Db 1401 ACTCTCGGTAAAGACGGCGAGTGTAAATCTTATTTGGTGGCAAAAGTGAAGAACGAGGTGTG 1460
QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTTTACGCGAGGCAAAAAATCACCATCAGCGAT 1080
Db 1461 ATTAGCGTAAATGGTGGCAGCATTTCTTTTACGCGAGGCAAAAAATCACCATCAGCGAT 1520
QY 1081 ATAATAAACCCCAACCATTTACTTTACAGCATTTCCGCGCTGAAATGAAGCGGTCAATCTG 1140
Db 1521 ATAATAAACCCCAACCATTTACTTTACAGCATTTCCGCGCTGAAATGAAGCGGTCAATCTG 1580
QY 1141 GCGGATATTTTTCGCAAAAGGCGGTAAACATTAATGTCGCTGCTGCACTATTTCGAAACCAA 1200
Db 1581 GCGGATATTTTTCGCAAAAGGCGGTAAACATTAATGTCGCTGCTGCACTATTTCGAAACCAA 1640
QY 1201 GGTAAACCTTCTGCTGATCTGTGAAGCAAAAGATAAAAGCGCAATATTTGCTTTCCGCC 1260
Db 1641 GGTAAACCTTCTGCTGATCTGTGAAGCAAAAGATAAAAGCGCAATATTTGCTTTCCGCC 1700
QY 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320
Db 1701 AAAGAGGTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1760
QY 1321 GGCAGCTGATGATTTACAGCGGATTAAGTCACATTTAAAAACAGGTGCACTTTATCGACCTT 1380
Db 1761 GGCAGCTGATGATTTACAGCGGATTAAGTCACATTTAAAAACAGGTGCACTTTATCGACCTT 1820
QY 1381 TCAGGTAAAGAGGAGGAGAACTTACCTTTGGCGGTGAGACCGCGGCGGAGGTAAAAAC 1440
Db 1821 TCAGGTAAAGAGGAGGAGAACTTACCTTTGGCGGTGAGACCGCGGCGGAGGTAAAAAC 1880
QY 1441 GGCATTCAATTTAGCAAGAAAAACCTCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1500

```



Db 1881 GGCAATTCATTAAGCAAGAAACCCCTTTAGAAAAGAGCTCAACCAATCAATGTATCAGGC 1940  
 QY 1501 AAAGAAAAGGCGGCGCTATTGTTGGGGCGGATATGCGTTAAATGAGCGCAATATT 1560  
 Db 1941 AAAGAAAAGGCGGCGCTATTGTTGGGGCGGATATGCGTTAAATGAGCGCAATATT 2000  
 QY 1561 AACGCTCAAGGTAGTGGTGATATCGCTAAACCCGGTGGTGGGAGACATCGGGCAT 1620  
 Db 2001 AACGCTCAAGGTAGTGGTGATATCGCTAAACCCGGTGGTGGGAGACATCGGGCAT 2060  
 QY 1621 TATTTATCCATTCAGACCAATGCAATTTGTTAAACCAAAAGAGTGGTGGTGGTGGTGGT 1680  
 Db 2061 TATTTATCCATTCAGACCAATGCAATTTGTTAAACCAAAAGAGTGGTGGTGGTGGTGGT 2120  
 QY 1681 GATGTACCAATTCGAGCAATGCAATTTGTTAAACCAAAAGAGTGGTGGTGGTGGTGGTGGT 1740  
 Db 2121 GATGTACCAATTCGAGCAATGCAATTTGTTAAACCAAAAGAGTGGTGGTGGTGGTGGTGGT 2180  
 QY 1741 CCAACAGCCCGGTGAAGCAAGCGCCCTTAAACCAAAAGAGTGGTGGTGGTGGTGGTGGTGGT 1800  
 Db 2181 CCAACAGCCCGGTGAAGCAAGCGCCCTTAAACCAAAAGAGTGGTGGTGGTGGTGGTGGTGGT 2240  
 QY 1801 ACCAATCAACTATTTCAATTTGTTAAACCAAAAGAGTGGTGGTGGTGGTGGTGGTGGTGGT 1860  
 Db 2241 ACCAATCAACTATTTCAATTTGTTAAACCAAAAGAGTGGTGGTGGTGGTGGTGGTGGTGGT 2300  
 QY 1861 AGAAAACCTTACGTTATAGCTCAATCAATCGAAGCAACCTCCCACTTAAATCTCCAT 1920  
 Db 2301 AGAAAACCTTACGTTATAGCTCAATCAATCGAAGCAACCTCCCACTTAAATCTCCAT 2360  
 QY 1921 AGTAAAGGTACGTTGCGGGAGCGTTGAGATTCATGATGAGATATTAATCTTAAAGCGGA 1980  
 Db 2361 AGTAAAGGTACGTTGCGGGAGCGTTGAGATTCATGATGAGATATTAATCTTAAAGCGGA 2420  
 QY 1981 AATTTAACCATTATTTCTGGCGGATGGTGTGATGTTCAATAAATATTTACGCTTGTATCAG 2040  
 Db 2421 AATTTAACCATTATTTCTGGCGGATGGTGTGATGTTCAATAAATATTTACGCTTGTATCAG 2480  
 QY 2041 GGTGTTTTTAAATATTAACCGCGCTTCCGTAGCTTTTGGAGTGGGAAATTAACAAAGCAGCG 2100  
 Db 2481 GGTGTTTTTAAATATTAACCGCGCTTCCGTAGCTTTTGGAGTGGGAAATTAACAAAGCAGCG 2540  
 QY 2101 GAGCGGCAATGCTAAATTTGTCGCCAGGCGACTGTAACCTTACAGGAGAGGGAATA 2160  
 Db 2541 GAGCGGCAATGCTAAATTTGTCGCCAGGCGACTGTAACCTTACAGGAGAGGGAATA 2600  
 QY 2161 GATTTCAAGGCTAACAACGTATCTTTAAACGGGAACGGGTAAGGCTGTAATATCAATTTCA 2220  
 Db 2601 GATTTCAAGGCTAACAACGTATCTTTAAACGGGAACGGGTAAGGCTGTAATATCAATTTCA 2660  
 QY 2221 TCAGTGAATATTTAAACCCCAATCTTTAGTGGCACAATTAACATATCTGGGAATATACA 2280  
 Db 2661 TCAGTGAATATTTAAACCCCAATCTTTAGTGGCACAATTAACATATCTGGGAATATACA 2720  
 QY 2281 ATTACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACCGCCATGATTCGCACTGG 2340  
 Db 2721 ATTACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACCGCCATGATTCGCACTGG 2780  
 QY 2341 AACGTCAGTCTTAAATCTAGACAGCGGCAAAATTTTACCTTTATTAATACATTTCA 2400  
 Db 2781 AACGTCAGTCTTAAATCTAGACAGCGGCAAAATTTTACCTTTATTAATACATTTCA 2840  
 QY 2401 AGCAATAGCAAGGCTTACACACAGTATAGAGCTCTGAGGGGTGAATTTTAAAGCGG 2460  
 Db 2841 AGCAATAGCAAGGCTTACACACAGTATAGAGCTCTGAGGGGTGAATTTTAAAGCGG 2900  
 QY 2461 GTAATGGCAACATGTCATTTCAATCTCAAGAGGAGCGGAAGTTAATTTCAAAATTAATA 2520  
 Db 2901 GTAATGGCAACATGTCATTTCAATCTCAAGAGGAGCGGAAGTTAATTTCAAAATTAATA 2960  
 QY 2521 CCAACAGAGAACATGACAAACAGCAAACTTTACCAATTCGTTTTTACGCAATATCACA 3020  
 Db 2961 CCAACAGAGAACATGACAAACAGCAAACTTTTACCAATTCGTTTTTACGCAATATCACA

QY 2581 GCCACTGTGGGGCTCTGTTTTTTTGTATATATATGCAACCAATCTGGCAGAGGGCT 2640  
 Db 3021 GCCACTGTGGGGCTCTGTTTTTTTGTATATATATGCAACCAATCTGGCAGAGGGCT 3080  
 QY 2641 GAGTTAAATCAGTCAATTAATATCTTAACGGCGCTAATTTTACCTTAAATTCOCAT 2700  
 Db 3081 GAGTTAAATCAGTCAATTAATATCTTAACGGCGCTAATTTTACCTTAAATTCOCAT 3140  
 QY 2701 GTTCGGCGGATGACGCTTTTAAATCAACAAAGACTTAAACCAATTAATGCAACCAATTC 2760  
 Db 3141 GTTCGGCGGATGACGCTTTTAAATCAACAAAGACTTAAACCAATTAATGCAACCAATTC 3200  
 QY 2761 AATTTCAAGCTCAGACAGAGCAAGATGATTTTATGCGGTGACGCAATGCCATC 2820  
 Db 3201 AATTTCAAGCTCAGACAGAGCAAGATGATTTTATGCGGTGACGCAATGCCATC 3260  
 QY 2821 AATTTCAAGCTCAGACAGAGCAAGATGATTTTATGCGGTGACGCAATGCCATC 2880  
 Db 3261 AATTTCAAGCTCAGACAGAGCAAGATGATTTTATGCGGTGACGCAATGCCATC 3320  
 QY 2881 AGCAGCAGCATTTACGGGGAATATTTACTATCGAAGAGCAGCAATGTTACGCTAGAAGCC 2940  
 Db 3321 AGCAGCAGCATTTACGGGGAATATTTACTATCGAAGAGCAGCAATGTTACGCTAGAAGCC 3380  
 QY 2941 AATAAGCGCCCTAATCAAGCAAAACATAGGGATAGAGTTTAAACTTGGCAGCTTGCCTC 3000  
 Db 3381 AATAAGCGCCCTAATCAAGCAAAACATAGGGATAGAGTTTAAACTTGGCAGCTTGCCTC 3440  
 QY 3001 GTTAAATGGAGTTTAAAGTTTAACTGGGGAATATGATTTAAAGCAATCTCACTATT 3060  
 Db 3441 GTTAAATGGAGTTTAAAGTTTAACTGGGGAATATGATTTAAAGCAATCTCACTATT 3500  
 QY 3061 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCTTAAATATATCACCAGCAATTTT 3120  
 Db 3501 TCAGAAAGCGGCACCTTTTAAAGGAAAGACTAGAGATACCTTAAATATATCACCAGCAATTTT 3560  
 QY 3121 ACCAATATGGCAGCTGCGGAAATTAATATACACAAAGAGTGGTAAACTTGGCAATGTT 3180  
 Db 3561 ACCAATATGGCAGCTGCGGAAATTAATATACACAAAGAGTGGTAAACTTGGCAATGTT 3620  
 QY 3181 ACCAATATGGCAGCTGCGGAAATTAATATACACAAAGAGTGGTAAACTTGGCAATGTT 3240  
 Db 3621 ACCAATATGGCAGCTGCGGAAATTAATATACACAAAGAGTGGTAAACTTGGCAATGTT 3680  
 QY 3241 GCGGAGATATTAACCAAAAGAGTGGTAAACTTAACTACAGACAGTAAATATATGATGCT 3300  
 Db 3681 GCGGAGATATTAACCAAAAGAGTGGTAAACTTAACTACAGACAGTAAATATATGATGCT 3740  
 QY 3301 GAAATCAAAATTTGGCGGCAATATCTCGCAAAAGAGGCAACCTCAGGATTTCTCCGAT 3360  
 Db 3741 GAAATCAAAATTTGGCGGCAATATCTCGCAAAAGAGGCAACCTCAGGATTTCTCCGAT 3800  
 QY 3361 AAAATTAATATPACCAACAGATTAACCAAAAGAGTGGTAAACTTAACTACAGACAGTAAATATGATGCT 3420  
 Db 3801 AAAATTAATATPACCAACAGATTAACCAAAAGAGTGGTAAACTTAACTACAGACAGTAAATATGATGCT 3860  
 QY 3421 TCAGATGCGACAGTAAATGCGCAACCTTAACTATTAACCAAGAGTGGTAAACTTAACTACAGACAGTAAATATGATGCT 3480  
 Db 3861 TCAGATGCGACAGTAAATGCGCAACCTTAACTATTAACCAAGAGTGGTAAACTTAACTACAGACAGTAAATATGATGCT 3920  
 QY 3481 GACCTAAAGTATTTTACAGGTTTCAATTAAGCAGAGATTTACAGCAGAGTGGTAACTTAACTACAGACAGTAAATATGATGCT 3540  
 Db 3921 GACCTAAAGTATTTTACAGGTTTCAATTAAGCAGAGATTTACAGCAGAGTGGTAACTTAACTACAGACAGTAAATATGATGCT 3980  
 QY 3541 ACTTATGGCAACAGTAAATGACGGTAAACAGCGGTGCGGAGCGCAACCAAGTAACTTAACTACAGACAGTAAATATGATGCT 3600  
 Db 3981 ACTTATGGCAACAGTAAATGACGGTAAACAGCGGTGCGGAGCGCAACCAAGTAACTTAACTACAGACAGTAAATATGATGCT 4040  
 QY 3601 AATGTTAAAGTTCACAAATCTCTGCTGACGGTCACTTAACTGACACTTAACTGACAAAGTGG 3660  
 Db 4041 AATGTTAAAGTTCACAAATCTCTGCTGACGGTCACTTAACTGACAAAGTGGTAACTGACAAAGTGG 4100

QY 3661 AAAACATCTGACGCAATGGCGGACGTGAAAGCAATAGCGACAAACGATACCGGCTTAAC 3720  
 Db 4101 AAAACATCTGACGCAATGGCGGACGTGAAAGCAATAGCGACAAACGATACCGGCTTAAC 4160  
 QY 3721 ATTACTGCAAAAATGTAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATAATC 3780  
 Db 4161 ATTACTGCAAAAATGTAGAGTAACAAAGATATTACTTCTCTCAAAACAGTAATAATC 4220  
 QY 3781 ACCGGCTGGAAAGTTACCAACAGGAGGCTCGACCAATTAACGCAAAATGGCAA 3840  
 Db 4221 ACCGGCTGGAAAGTTACCAACAGGAGGCTCGACCAATTAACGCAAAATGGCAA 4280  
 QY 3841 GCAAGTATTACCAACAGGAGGCTCGACCAATTAACGCAAAATGGCAA 3900  
 Db 4281 GCAAGTATTACCAACAGGAGGCTCGACCAATTAACGCAAAATGGCAA 4340  
 QY 3901 GTTAGCGGAGTGGGATTAACCACTAAATCCGGCTCAAAATTAAGCGAAATCGGGT 3960  
 Db 4341 GTTAGCGGAGTGGGATTAACCACTAAATCCGGCTCAAAATTAAGCGAAATCGGGT 4000  
 QY 3961 GAGGCTAAATGTAACAGTGCACAGGTACAATGGCGGTACAATTTCCGGTAATACGGTA 4020  
 Db 4401 GAGGCTAAATGTAACAGTGCACAGGTACAATGGCGGTACAATTTCCGGTAATACGGTA 4080  
 QY 4021 AATGTTAGCGCAACGCTGGCGATTTACAGTTGGGATGGCGAGAAATTAATGCGACA 4080  
 Db 4461 AATGTTAGCGCAACGCTGGCGATTTACAGTTGGGATGGCGAGAAATTAATGCGACA 4140  
 QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGAGTAACCTTGACTACTGAAGCGGTTCTAGC 4140  
 Db 4521 GAAGGAGCTGCAACCTTAACCGCAACAGGAGTAACCTTGACTACTGAAGCGGTTCTAGC 4180  
 QY 4141 ATCACTTCACTAAGGTCAGGTAGACTCTTGCGCTCAGAGTGCAGCAAAATTAATGCGACA 4200  
 Db 4581 ATCACTTCACTAAGGTCAGGTAGACTCTTGCGCTCAGAGTGCAGCAAAATTAATGCGACA 4260  
 QY 4201 ATTAATGCTGTAATGTGACATTAATTAACCTTACAGGCACTTAACCAACCGTGGCAGGCTCG 4260  
 Db 4641 ATTAATGCTGTAATGTGACATTAATTAACCTTACAGGCACTTAACCAACCGTGGCAGGCTCG 4320  
 QY 4261 GATATTAACCAACCGTGGCAGGCTTAATTAACCGCAACAGGAGTAACTGTAAGCTTAATGGT 4320  
 Db 4701 GATATTAACCAACCGTGGCAGGCTTAATTAACCGCAACAGGAGTAACTGTAAGCTTAATGGT 4380  
 QY 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCGTAGTGTG 4380  
 Db 4761 GATGATCAGGTGATAGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCGTAGTGTG 4440  
 QY 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGCGGATTTAAACACAGTAATGGCTTA 4440  
 Db 4821 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGCGGATTTAAACACAGTAATGGCTTA 4480  
 QY 4441 AATATCATTTTCAAGAGTGGTAGAAGCACTGCGCTTAAGAGCAAGGAAATGGAGGTG 4500  
 Db 4881 AATATCATTTTCAAGAGTGGTAGAAGCACTGCGCTTAAGAGCAAGGAAATGGAGGTG 4560  
 QY 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAGCAAGTAAATGAAGCAAGCGCTTCCT 4560  
 Db 4941 AATATATCCAGCAGGTGTAGCAAGTGTAGAGCAAGTAAATGAAGCAAGCGCTTCCT 5000  
 QY 4561 GAAAAGTAAAGATTTATCTGATGAAGCAAGTAAATGAAGCAAGCGCTTCCT 4620  
 Db 5001 GAAAAGTAAAGATTTATCTGATGAAGCAAGTAAATGAAGCAAGCGCTTCCT 5060  
 QY 4621 GCTGATGTTTGTAGGCAAAATAATACAAATTAACAGTCAATACAGCAATGAATTTACA 4680  
 Db 5061 GCTGATGTTTGTAGGCAAAATAATACAAATTAACAGTCAATACAGCAATGAATTTACA 5120  
 QY 4681 ACCAGCCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 4740  
 Db 5121 ACCAGCCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 5180  
 QY 4741 GGCGCAGGAGTATGTACCAATGTGCTGACGATGGACGCGCTAGTCAAGTAATGACAAG 4800

Db 5181 GGCGCAGGATGTACCAATGTGCTGAGCATGGACCGGTAGTCAATTTGACAAG 5240  
 QY 4801 GTAGATTTTCCTGCAATGAAGTCAATTTATTTTCGTATTATTTCTGTTGGGTAA 4860  
 Db 5241 GTAGATTTTCCTGCAATGAAGTCAATTTATTTTCGTATTATTTCTGTTGGGTAA 5300  
 QY 4861 GTTCAGTACGGGCTTTACCCATCTTCTAAATAATACGAGAAATCAATAAGATTTTT 4920  
 Db 5301 GTTCAGTACGGGCTTTACCCATCTTCTAAATAATACGAGAAATCAATAAGATTTTT 5360  
 QY 4921 AACAGTTATTATTAG 4937  
 Db 5361 AACAGTTATTATTAG 5377

RESULT 14  
 HIU08876  
 LOCUS  
 DEFINITION  
 Haemophilus influenzae 12 adhesin (hmw1A) gene and putative accessory processing proteins (hmw1B) and (hmw1C) genes, complete cds.  
 U08876 M84616  
 U08876.1 GI:475770  
 SOURCE  
 Haemophilus influenzae.  
 Haemophilus influenzae  
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 Haemophilus.  
 1 (bases 1 to 5116)  
 Barenkamp, S.J. and Leininger, E.  
 Cloning, expression, and DNA sequence analysis of genes encoding non-pyruvate Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis  
 Infect. Immun. 60 (4), 1302-1313 (1992)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 1548058  
 2 (bases 5117 to 9221)  
 Barenkamp, S.J. and St Geme, J.W. III.  
 Genes encoding high-molecular-weight adhesion proteins of non-pyruvate Haemophilus influenzae are part of gene clusters  
 Infect. Immun. 62 (8), 3320-3328 (1994)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 8039903  
 3 (bases 1 to 9221)  
 Barenkamp, S.J.  
 Direct Submission  
 Submitted (14-APR-1994) Stephen J. Barenkamp, Department of Pediatrics, St. Louis University School of Medicine, Cardinal Glennon Children's Hospital, 1465 South Grand Boulevard, St. Louis, MO 63104-1095, USA  
 On Apr 29, 1994 this sequence version replaced gi:148902.

## source

Location/Qualifiers  
 1. 9221  
 /organism="Haemophilus influenzae"  
 /strain="12"  
 /db\_xref="taxon:727"  
 /clone\_lib="Lambda EMBL3 genomic library of strain 12"  
 129..240  
 /note="direct"  
 /citation=[1]  
 /rpt\_type=tandem  
 351..4961  
 /gene="hmw1A"  
 351..4961  
 /gene="hmw1A"  
 /note="surface-associated protein"  
 /citation=[1]  
 /codon\_start=1  
 /transl\_table=11  
 /product="adhesin"  
 /protein\_id="AAA20527.1"

## repeat\_region

## gene

## CDS



|||||  
Db 1000 AAGTGAACACGAGGCTGATAGCGTAAATGGTGGCAGCATTTCTTTACTCGCAGGCG 1059  
QY 1061 AAAAAATCACCATCAGCGATATATAAACCACCAACCATTTACTTACAGCATTTGCCGCGCGTG 1120  
Db 1060 AAAAAATCACCATCAGCGATATATAAACCACCAACCATTTACTTACAGCATTTGCCGCGCGTG 1119  
QY 1121 AAAAAATCAGCGTCAATCTGGCGGATATATTTTGGCAAGGCGGTAAACATTAATCTCCGCTG 1180  
Db 1120 AAAAAATCAGCGTCAATCTGGCGGATATATTTTGGCAAGGCGGTAAACATTAATCTCCGCTG 1179  
QY 1181 CTGCCACTATTCGAAACCAAGGTAACTTTCTGCTGATTTCTGTAAGCAAAAGATATAAGCG 1240  
Db 1180 CTGCCACTATTCGAAACCAAGGTAACTTTCTGCTGATTTCTGTAAGCAAAAGATATAAGCG 1239  
QY 1241 GCAATATTTGTTCTTTCCGCCAAAGAGGTGAAGCGGAATTTGGCGGTGAATTTCCGCTC 1300  
Db 1240 GCAATATTTGTTCTTTCCGCCAAAGAGGTGAAGCGGAATTTGGCGGTGAATTTCCGCTC 1299  
QY 1301 AAAAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTAACAGCGCATTAAGTCACATTAATAA 1360  
Db 1300 AAAAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTAACAGCGCATTAAGTCACATTAATAA 1359  
QY 1361 CAGGTGAGTATTCAGCTTTTCAAGTAAAGAGGGGAGAACTTACCTTGGCGGTGAGC 1420  
Db 1360 CAGGTGAGTATTCAGCTTTTCAAGTAAAGAGGGGAGAACTTACCTTGGCGGTGAGC 1419  
QY 1421 AGCGCGGCAAGGTAAACAGCGCATTTCAATTAAGCAAGAAACCTTTAGAAAAAGGCT 1480  
Db 1420 AGCGCGGCAAGGTAAACAGCGCATTTCAATTAAGCAAGAAACCTTTAGAAAAAGGCT 1479  
QY 1481 CAACCATCAATGATCAGCAAGAAAGAGCGGCGCTATTTGTTGGGGGATATTTG 1540  
Db 1480 CAACCATCAATGATCAGCAAGAAAGAGCGGCGCTATTTGTTGGGGGATATTTG 1539  
QY 1541 CGTTAATTTAGCGCAATTAATTAACGCTCAAGGTAGTGTGATGATCGCTTAACACCGTGGTT 1600  
Db 1540 CGTTAATTTAGCGCAATTAATTAACGCTCAAGGTAGTGTGATGATCGCTTAACACCGTGGTT 1599  
QY 1601 TTGTGGAGACATCGGGCATATTTATTCATTCACAGCAATGCAATTTGTTAAACAAAG 1660  
Db 1600 TTGTGGAGACATCGGGCATATTTATTCATTCACAGCAATGCAATTTGTTAAACAAAG 1659  
QY 1661 AGTGGTTGCTAGACCTGATGATTAACATTAACATTAACATTAACATTAACATTAACATTAAC 1720  
Db 1660 AGTGGTTGCTAGACCTGATGATTAACATTAACATTAACATTAACATTAACATTAACATTAAC 1719  
QY 1721 CCGGTATTAATGATGATTTCCCAACAGGCGCGGTGAAGCAAGCCCTTGAACCAAGCTTAC 1780  
Db 1720 CTTCAAGACGATGATTAACAGGCGGTGAAGCAAGCCCTTGAACCAAGCTTAC 1779  
QY 1781 CGCACTCAAAACACGCTTAACCAATCAACTATTTCAAAATTTCTGAAAAACGCCCTGGA 1840  
Db 1780 AAGAA---AAGACAACATTAACCAACACTCTTGAGAGTATTAACCAAAAGGTACCT 1836  
QY 1841 CAATGAATATACGGCATCAAGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCA 1900  
Db 1837 TTGTTTAACTACATGCTAATCAAGCGCATCTATGTCATAGCTCAATTAATTT---ATCA 1893  
QY 1901 ACTCCACTTAATTTCTCCTAGTAAAGGTAGGCGGTGCGGAGCGCTTACAGTTGATGGAG 1960  
Db 1894 ATGGCAGCTTAACCTTTGAGGTAGGCGGTGCGGAGCGGTGCGGAGCTTGAATTAACAGC 1953  
QY 1961 ATATTAC-----TTCTAAAGCGGAAATTTAAACCATTTTAACTTCTGCGGATGGG 2008  
Db 1954 ATATTACCGCGGTGATGATACAGAGGTGCAAACTTAACTTAACTTAACTTAACTTAACTTAA 2013  
QY 2009 TTGATGTTCAAAAAATATATAGCTTGTAGTACAGGTTTAAATATTAACCGCGCTTCCG 2068  
Db 2014 TTGATGTTCAAAAAATATCTCACTCGGGCGCAAGGTAAACATAAACAATTAACAGCTAAG 2073  
QY 2069 TAGCTTTTCAAGGTGAATATACAAAGCACCGCAGCGGCAAAATGCTAAATTTCTCCGCC 2128  
|||||

Db 2074 AAGATATCGCTTTTGAGAAAGGAGCAACCAAGTC-----ATTACAGTC 2118  
QY 2129 AGGCACTCTTACCATTACAGGAGAGGAAAAAGATTTTCAGGGCTAACCAAGCTATCTTTAA 2188  
Db 2119 AAGGCACTATTACC---TCAGGCAATCAAAAAGGTTTATAGATTTAAATACTCTCTCTAA 2175  
QY 2189 ACGGAACGGTAAAGCTCTGAATATCATTTTCATCAGTGAATAA-----TTTAAACC 2239  
Db 2176 ACGGCACTGGCAGCGGACTGCAATTCACCAATAAAGAACCAATAAATACGCTATCAAA 2235  
QY 2240 ACAATCTTAGTGGCACATTAACATATCTGGGAATATAACAAT---TAACCAACTACGA 2296  
Db 2236 ATAATTTGAAGGAGCTTAATAATTTTTCAGGAAAGTGAACATCTCAATGGTTTTACCTA 2295  
QY 2297 GAAAGACACCTCGTATTTGGCAACACAGCCATGATTCGCACCTGCAAGCTCAGTGTCTTTA 2356  
Db 2296 AAAATGAAGTGGATATGATAAATTCAGAGGACGCTTACTTGAATTTTAACTCTCTTA 2355  
QY 2357 ATCT---AGAGACAGCGGCNAATTTTACCTTTTATTAATACATTTCAAGCAATAGCAAG 2413  
Db 2356 ATGTTCCGAGAGTGGGAGTTTAACTCTCACTATTGACTCCAGAGAAAGCGATAGTCAG 2415  
QY 2414 GCTTAACACACAGTATAGAAGCTCTGCAAGGGTGAATTTTAAACGGCTTAATGGCAACA 2473  
Db 2416 GCACCTTACCAGCGCTTATAATTTAAACGGTATATCATTTCAACAAGACACT---ACCT 2472  
QY 2474 TGTCATTCAATCTCAAGAGGAGCGAAAGTTTAAATTTCAAAATTAACCAACAGCAACA 2533  
Db 2473 TTAATGTTGAAGCAATGCAAGAGTCAACTTTGACATCAAGGCAACCAATAGGTAATA 2532  
QY 2534 TGAACACAGCAACCTTTTACCAATTCGGTTTATAGCAATATCAAGCACTGTTGGGG 2593  
Db 2533 AGTATCTAGTTTGAATTTACGCATCATTTAATGCAACATTTTCAAGTTTCGGGAGGGGA 2592  
QY 2594 GCTCTGTTT-----TTTTTGATATATATGCAACCACTTCTGCGAGAGGGCTGAGTTAA 2647  
Db 2593 GTGTTGATTTTCAACTCTCGCTCATCTTAACTGCAACCCCGGTGAGTTATAA 2652  
QY 2648 AATGAGTGAATTAATCTCTAAGCGGGCTAAATTTTACCTTTAAATTTCCCATTTCA 2767  
Db 2653 ATTCTAAATACTTTAATGTTTCAACAGGCTAAGTTTAAAGATTTTAACTTCAGGCTCA 2712  
QY 2708 GCGATGACGCTTTTAAATCAACAAAGACTTAACCAATTAACCAACCAATTTCAATTTCA 2767  
Db 2713 CAAAACTGCTTCTCAATAGAGAAAGATTTTAACTTTAAAGTCCACC-----GGAGCA 2766  
QY 2768 GCTCAGACAGCAAGAAAGATTTTATGCGGTGACGCAAGTCCATGCCATCAATCA 2827  
Db 2767 ACATAACACTTTTGCAGAGTTGAGGCAACGATGGAATGTTGTTAAAGGCAATTTAGCCA 2826  
QY 2828 CCTACAACATATCCATTTGCGCGGTAAATGTCACCTTTGGTGGCAAAACTCAAGCAGCA 2887  
Db 2827 AAAAAACATACCTTTGAAGGAGTAAACATCACCTTTGGCTCAGAAAGCGCTAACAG 2886  
QY 2888 GCATTACGGGAATATTACTATCGAGAAAGCAAGCAATTTAGCTAGAGAGCAATAACG 2947  
Db 2887 AAATCGAAGGCAATGTTACTATCAATTAACACGCTAACGCTACTTTATCGTTGCGATT 2946  
QY 2948 CCCCTAATCAGCAAAACATAAGGATAGAGTTATAAACTTTGCGACCTTCTGCTGTTAATG 3007  
Db 2947 TTGCAACCACTCAAAA-----ACCTTTAACTTTAAAAAGATGTCAATTAATAGCG 3000  
QY 3008 GGAGTTTAACTTTAACTGGCGAAATGCAATATTAAGGCAATCTCACTATTTTACAGAA 3067  
Db 3001 GCAACCTTACCGCTGGAGCAATATTTGCAATATAGCGGAAATCTTACCGTTGAAGTA 3060  
QY 3068 GCGCCACTTTTAAAGAAAGATAGAGATACCTTAATATCACCGCAATTTTACCATA 3127  
Db 3061 ACGCTAATTTCAAGCTATCAAAATTTTCACTTTTAAATGAGGCGCTTGTGTTGACAACA 3120  
QY 3128 ATGGCACTGCCAAATTAATATAACAGAGGAGTGAATAAACTTTGCAATTTTACCAATG 3187  
Db 3121 AAGGCAATTCAAATATTTCCTATTGCCAAGAGGGGCTCGCTTTTAAAGACATTTGATAAT 3180





TITLE High molecular weight surface proteins of non-typeable haemophilus  
JOURNAL Patent: US 5869065-A 1 09-FEB-1999;  
FEATURES Location/Qualifiers  
source l. 5116

BASE COUNT 1766 a 1008 c 1037 g 1305 t  
ORIGIN

Query Match 54.3%; Score 2680.2; DB 6; Length 5116;  
Best Local Similarity 73.5%; Pred. No. 0;  
Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps 16;

QY 41 ATGACAAACAAATATACAAACCTTTTTCAGTCTATATGCAATATATTTAAAAAAT 100  
|||||  
Db 41 ATGACAAACAAATATACAAACCTTTTTCAGTCTATATGCAATATATTTAAAAAAT 99  
|||||  
QY 101 AGTATAATCCGCCATATAAAATGGTATAATCTTTTCATCTTTTCAATCTTTTCATC 160  
|||||  
Db 100 AGTATAATCCGCCATATAAAATGGTATAATCTTTTCATCTTTTCAATCTTTTCATC 159  
|||||  
QY 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 220  
|||||  
Db 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 219  
|||||  
QY 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 280  
|||||  
Db 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 279  
|||||  
QY 281 GAATGAAGAGGAGCTGACGACGCAATATGATAAGTAATTTTCACTTAACCT 340  
|||||  
Db 280 GAATGAAGAGGAGCTGACGACGCAATATGATAAGTAATTTTCACTTAACCT 339  
|||||  
QY 341 TAGGAGAAATATGAACAGATATATCGTCTCAAAATTCAGCAAGGCGCTGATCTTTGG 400  
|||||  
Db 340 TAGGAGAAATATGAACAGCTATATCGTCTCAAAATTCAGCAAGGCGCTGATCTTTGG 399  
|||||  
QY 401 TTGCTGTGTCTGAATTTGGACGGGTTGTGACCATTCACAGAAAGGCTTCGCGTATG 460  
|||||  
Db 400 TTGCTGTGTCTGAATTTGGACGGGTTGTGACCATTCACAGAAAGGCTTCGCGTATG 459  
|||||  
QY 461 TTACTATCTTTTAGGTGTAACCACTTTAGGTTTAAAGCCACTTTCCGCTATGTACTATCT 520  
|||||  
Db 460 CTGCTCGCATGAAGTGGCTCACTTAGGTTTAAAGCCACTTTCCGCTATGTACTATCT 519  
|||||  
QY 521 TAGGTGTAACATCTATCCCAATCTGTTTTCAGCAAGCGGCTTACAGGAATGATGTAG 580  
|||||  
Db 520 TAGGTGTAACATCTATCCCAATCTGTTTTCAGCAAGCGGCTTACAGGAATGATGTAG 579  
|||||  
QY 581 TACAGGCACAGCCACTATGCAAGTAGATGTAATAAACCAATATCCGCAACAGTGTG 640  
|||||  
Db 580 TACAGGCACAGCCACTATGCAAGTAGATGTAATAAACCAATATCCGCAACAGTGTG 639  
|||||  
QY 641 ACGCTATCATTAATTTGAAACAATTTAACATCGACCAAAATGAAATGGTGGCTTTTAC 700  
|||||  
Db 640 ACGTATCATTAATTTGAAACAATTTAACATCGACCAAAATGAAATGGTGGCTTTTAC 699  
|||||  
QY 701 AAGAAACACAACTCCGCGTATTCACCGTGTACATCTAACCAAAATCCCAATTA 760  
|||||  
Db 700 AAGAAACACAACTCCGCGTATTCACCGTGTACATCTAACCAAAATCCCAATTA 759  
|||||  
QY 761 AAGGATTTTATGATCTTAACGACAGTCTTTTATCAACCCAAATGATATCACAATAG 820  
|||||  
Db 760 AAGGATTTTATGATCTTAACGACAGTCTTTTATCAACCCAAATGATATCACAATAG 819  
|||||  
QY 821 GTAAGACGAATTTTAACTAATGGCTTTTACGGCTTCTACGGCTAGACATTTCTAACG 880  
|||||  
Db 820 GTAAGACGAATTTTAACTAATGGCTTTTACGGCTTCTACGGCTAGACATTTCTAACG 879  
|||||  
QY 881 AAACATCAAGGCGCTAATTTTCACTTCGAGCAAAACCAAGATAAAGCGCTCGGTGAA 940  
|||||  
Db 880 AAACATCAAGGCGCTAATTTTCACTTCGAGCAAAACCAAGATAAAGCGCTCGGTGAA 939  
|||||  
QY 941 TTGTGAATCACGGTTTAAATTAATCTGCGGTAAGACGGCAGTGAATCTTATTGGTGCA 1000  
|||||

Db 940 TTGTGAATCACGGTTTAAATTAATCTGCGTAAAGACGGCAGTGAATCTTATTGTGGCA 999  
|||||  
QY 1001 AAGTGAATAAACAGGCTGTGATTAGCGTAAATGGTGGAGCATTCTTTACTCGCAGGC 1060  
|||||  
Db 1000 AAGTGAATAAACAGGCTGTGATTAGCGTAAATGGTGGAGCATTCTTTACTCGCAGGC 1059  
|||||  
QY 1061 AAAAAATCACCATCAGCATATATAAACCCCAACCATTTACTTACAGCATTTGCCGGCTG 1120  
|||||  
Db 1060 AAAAAATCACCATCAGCATATATAAACCCCAACCATTTACTTACAGCATTTGCCGGCTG 1119  
|||||  
QY 1121 AAAATGAAGCGTCAATCTGGCGGATATTTTGGCAAGGCGGTAACTTAATGTCCGTG 1180  
|||||  
Db 1120 AAAATGAAGCGTCAATCTGGCGGATATTTTGGCAAGGCGGTAACTTAATGTCCGTG 1179  
|||||  
QY 1181 CTGCGACTATTTCGAACCAAGTAACTTTCTGCTGATTCTGTAAAGCAAGATATAAGCG 1240  
|||||  
Db 1180 CTGCGACTATTTCGAACCAAGTAACTTTCTGCTGATTCTGTAAAGCAAGATATAAGCG 1239  
|||||  
QY 1241 GCAATATTGTTCTTTCCGCCAAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1300  
|||||  
Db 1240 GCAATATTGTTCTTTCCGCCAAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1299  
|||||  
QY 1301 AAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTACAGGCGATAAAGTCAACATTAATA 1360  
|||||  
Db 1300 AAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTACAGGCGATAAAGTCAACATTAATA 1359  
|||||  
QY 1361 CAGCTGAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1420  
|||||  
Db 1360 CAGCTGAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1419  
|||||  
QY 1421 AGCGCGCAAGGTAAAGCGGCAATTTCAATTTAGCAAGAAACCTCTTTAGAAAGAGGCT 1480  
|||||  
Db 1420 AGCGCGCAAGGTAAAGCGGCAATTTCAATTTAGCAAGAAACCTCTTTAGAAAGAGGCT 1479  
|||||  
QY 1481 CAACCATCAATGATCAGGCAAGAAAGAGCGGCGCTATTGTTGGCGGCTATTTG 1540  
|||||  
Db 1480 CAACCATCAATGATCAGGCAAGAAAGAGCGGCGCTATTGTTGGCGGCTATTTG 1539  
|||||  
QY 1541 CGTTAATTGACGCAATATTAACGCTCAAGGTAGTGGTATATCGCTTAAACCGGTGGTT 1600  
|||||  
Db 1540 CGTTAATTGACGCAATATTAACGCTCAAGGTAGTGGTATATCGCTTAAACCGGTGGTT 1599  
|||||  
QY 1601 TTGTGAGACATCGGGCATTTATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAG 1660  
|||||  
Db 1600 TTGTGAGACATCGGGCATTTATTTATTTATCAAGACCAATGCAATTTGTTAGCCAAAG 1659  
|||||  
QY 1661 AGTGGTTGCTAGACCTGATGATGTAACAATTTGAAGCCGAGACCCCTTTCGAATAA 1720  
|||||  
Db 1660 AGTGGTTGTTAGACCGGATAATGATCTATTATGACAGAAACAGCAGGACGCAATA 1719  
|||||  
QY 1721 CCGGTATAAATGATGAATTTCCCAACAGGCGCACGCTCAAGCAAGCGCACCTAAAAAATA 1780  
|||||  
Db 1720 CTTTCAAGACCATGATGATACAGGGATCGGGAATAGTCCAGCACCCCAACCAAGAA 1779  
|||||  
QY 1781 GCGAATCAAAACACGCTTAACCAATATTAACATTTATCTGAAACAGCGCTGGA 1840  
|||||  
Db 1780 AAGAA---AAGCAACATTAACAAACACAACTCTTTGAGAGTATCTAAAGAGTACCT 1836  
|||||  
QY 1841 CAATGAATATAACGCAATCAAGAAACTTACCGTTAATAGTCAATCAACATTCGGAAGA 1900  
|||||  
Db 1837 TTGTTTAACTACTCTATCAACCCATCTATGTCATTAAGTCTCAATTAATTT---ATCCA 1893  
|||||  
QY 1901 ACTCCACTTAAATTTCTCCATAGTAAAGGTGCGGTGGCGAGGCTTCAGATGATGAG 1960  
|||||  
Db 1894 ATGGCAGCTTAATCTTTGGAGTGAGGCTCGGAGCGGTGGCGGCTTGAGATTAACAACG 1953  
|||||  
QY 1961 ATATTAC-----TTCTAAAGCGGGAATTTAACCAATTTATTTCTGGCGGATGGG 2008  
|||||  
Db 1954 ATATTACCGCGGTGATGATACAGAGGTGCAAACTTTAACTTTTACTCAGCGGCTGGG 2013  
|||||  
QY 2009 TTGATGTTTCAAAAAATTTACGCTTGTATCAGGCTTTTAAATATTACCGGCTCGG 2068  
|||||



Db 2014 TTGATGTTTCAAAAATATCTCACTCGGGGCGCAAGGTAAACATTAACACTTACAGCTAAAC 2073  
 Qy 2069 TAGCTTTTGAAGGTGGAATTAACAAAGCACGCGCGCGGCAAAATGCTAAAAATTCGCGCC 2128  
 Db 2074 AAGATATCGCTTTTGAGAAAGGAAGCAACCAAGTC-----ATTACAGGTC 2118  
 Qy 2129 AGGCGACTGTACCAATTTACAGGAGAGGGAAGAAAGATTTTCAGGGCTAAACACGCTATCTTTAA 2188  
 Db 2119 AAGGCACTATTACC---TCAGGCAATCAAAAAGGTTTGTAGATTAAATATGCTCTCTAA 2175  
 Qy 2189 AGGCAACGGGTAAAGGTGTGAATATCAATTCATCAGTCAATAA-----TTTAAACCC 2239  
 Db 2176 AGGCACTGGCAGCGGACTGCAATTCACCACTAAAGAACCAATAAAATACGCTATCAACA 2235  
 Qy 2240 ACAATCTTAGTGGCACAAATTAACATATCTGGGAATATTAACAT---TAACCAAACTACGA 2296  
 Db 2236 ATAAATTTGAAGGGACTTTAAATATTTCAAGGAAAGTGAACATCTCAATGGTTTACCTA 2295  
 Qy 2297 GAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGGAAAGTCAGTCTCTTA 2356  
 Db 2296 AAAATGAAAGTGGATATGATAAATTAAGAGGACGCTTACTTGGAAATTTAACTCTCTAA 2355  
 Qy 2357 ATCT---AGAGACAGCGCAAAATTTTACCTTTTATTAATATACATTTCAAGCAATACGAAG 2413  
 Db 2356 ATGTTTCCGAGAGTGGCGAGTTTAACTTACCTTACCTTACCTTCCAGAGGAAGCGATAGTGAG 2415  
 Qy 2414 GCTTAAACAACACAGTATAGAGCTCTGACAGGGGTGAATTTTAAACGGCTTAAATGGCAACA 2473  
 Db 2416 GCACACTTACCCAGCTTTATATTTTAAACGGTATATCAATTCACAAAGACACT---ACCT 2472  
 Qy 2474 TGCTATCAATCTCAAGAGGAGCGAAAGTTAAATTTCAATTTAAACCAACAGCAACA 2533  
 Db 2473 TTAATGTTGAACAAATGCAAGAGTCAACTTGCACATCAAGGCACCAATAGGATAAATA 2532  
 Qy 2534 TGAACACAGCAAACTTTACCAATTTTCGGTTTAAAGCAATATCAACCACTGCTGGG 2593  
 Db 2533 AGTATCTTAGTGAATAGCGCATATTAATGGAACATTTTCACTTTCGGGAGGGGA 2592  
 Qy 2594 GCTCTGPTT-----TTTTTGATATATATGCCCAACCATTTCTGGCAGAGGGCTGAGTTAA 2647  
 Db 2593 GTGTGATTTTACACTTCTCGCTCTACCTCTAACCTTCCAAACCCCGGTGATGTTATAA 2652  
 Qy 2648 AAATGAGTGAATTAATCTCTAAACGGCGCTAATTTTACCTTTAAATTTCCCATGTTCCGG 2707  
 Db 2653 ATTCTAAATACCTTTAATGTTTCAACAGGGTCAAGTTTAAAGTTTAAACCTTCAAGGCTCAA 2712  
 Qy 2708 GCGATGACGCTTTTAAATCAACAAAGACTTAACCAATAATGCAACCAATTTCAAAATTTCA 2767  
 Db 2713 CAAAACTGGCTTCTCAATAGAGAAAGATTTAACTTTTAAATGCCACC-----GGAGGCA 2766  
 Qy 2768 GCCTCAGACAGACGAAAGATGATTTTATGACGGGTACGACGCAATGCGCATCAATTCAA 2827  
 Db 2767 ACATAACACTTTTCAAGTTGAAGCCCGCATGGAATGATTTGTTAAAGCATTTGAGCCA 2826  
 Qy 2828 CCTCACAATATCACTTCTGGCGGTAAATGTCACCTTTGGTGACAAACCAACCAAGCAGCA 2887  
 Db 2827 AAAAAACATTAACCTTTGAAGAGGTACATCACTTTGGCTCCAGGAAGCCGTACAG 2886  
 Qy 2888 GCATTACGGGGAATATTACTATCGAGAAAGCAGCAATGTTACGCTAGAGCCCAATTAACG 2947  
 Db 2887 AAATGGAAGGCAATGTTTACTATCAATAACAGCTTAACGTCATCTTATCGGTTCCGGATT 2946  
 Qy 2948 CCCCTAATCAGCAAAACATAAGGATAGATTTATAAACTTTGGCAGCTTGCCTGCTTAATG 3007  
 Db 2947 TTGCAACCACTAAAA-----ACCTTTAACTTTAAAAAGATGTCTATCTAATTAAGCG 3000  
 Qy 3008 GGAGTTTAACTTAACTGCGGAAATGAGATATTAAAGCAATCTCACTATTTCAGAAA 3067  
 Db 3001 GCACCTTTACCGCTGGAGCAATATTGCAATATAGCCGGAATCTTTACCGCTTGAAGTA 3060  
 Qy 3068 GCGCCACTTTAAAGGAAGACTAGATACCTTAATATACCGGCAATTTTACCATA 3127  
 Db 3061 ACGCTAATTTCAAGCTATCACAAAATTTCACTTTTAAATGAGCGGCTTTGTTGACACA 3120

Qy 3128 ATGCACTGCGGAATTAATATAACACAGGAGTGGTAAACCTTGGCAATGTTTACCAATG 3187  
 Db 3121 AAGCAATTCAAATATTTCCNTTGGCAAGGAGGGCTCGCTTTAAAGACATTTGATAAT 3180  
 Qy 3188 ATGCTGATTTAAACATTTACCCTCACCGTAAACGCAACCAAGCAATCATCGCGGAG 3247  
 Db 3181 CCAAGAAATTTAAGCATCATCCACCAACTCCAGCTCCACTTACCCGACTATTATTAAGCGCA 3240  
 Qy 3248 ATATAATCAACAAAAAGGAGCTTAAATATACAGACATAATATGATGCTGGAATCC 3307  
 Db 3241 ATATAACCAATAAAACGGTATTTAAATATACGAACGAAGGTAGTACTGCAAAATGC 3300  
 Qy 3308 AAATTTGGCGGCAATATCTCGCAAAAAGAGCAACCTCACGATTTCTTCGATAAAATTA 3367  
 Db 3301 AAATTTGGCGGCAATCTCTCGCAAAAAGAGGTATCTCACGATTTCTTCTGCAAAAATCA 3360  
 Qy 3368 ATATCACCAACACATTAACATCAAAAAGGGTATTTGATGGAGAGGACTAGTTCAGATG 3427  
 Db 3361 ATATTACCAACACATTAACATCAAGCAGGTGTTGATGGGAGAAATTCGGATTCAGACG 3420  
 Qy 3428 CGAAGTAAATGCCAACCCTAACTATTAACCAAAAGAAATTCGAAATTCACAGAACCTAA 3487  
 Db 3421 CGCAACAAATGCCAATCTAACCATTAACCAAAAGAAATTCGAAATTAACGCAAGACCTAA 3480  
 Qy 3488 GTATTTCAAGTTTCAATTAACAGCAGATTTACAGCAAGATGGTAGAGATTTAACTATTG 3547  
 Db 3481 ATATTTCAAGTTTCAATTAACAGCAGATTTACGCTAAAGATGGTAGTATTTAACTATTG 3540  
 Qy 3548 GCAACAGTAATGACGGTTAACAGCGGTGCGGAAGCCAAACAGTAACCTTTTAAACAATGTA 3607  
 Db 3541 GTAACACCAATATAGTCTGA---TGGTACTAATGCGCAAAAAGTAACCTTTTACCAAGTTA 3597  
 Qy 3608 AAGATTCAAAAATCTCTCTGACGGTCAAAATGTGACACTAAATAGCAAAAGTGAACAT 3667  
 Db 3598 AAGATTCAAAAATCTCTCTGACGGTCAAAAGGTGACACTACACAGCAAGTGAACAT 3657  
 Qy 3668 CTAGCAGCAATGGCGGAGCTGAAAGCAATAGCGCAACAGTACCGGCTTAACTATTACTG 3727  
 Db 3658 CCGGTAGTAATTAACCAACTTGAAGATAGCAGTGAACAATAATGCGGCTTAACTATCGATG 3717  
 Qy 3728 CAAAAATGTACAAGTAAACAAAGATATTACTTCTCTCAAAACAGTAATATCAACCGGT 3787  
 Db 3718 CAAAAATGTACAGTAAACAAATATTACTTCTCAAAAGCAGTGAAGCATCTCTGCGA 3777  
 Qy 3788 C---GGAAAGTTTACCACACAGCAGCTCGACCATTAACGCAACCAATTAACGCAACCTG 3839  
 Db 3778 CAAGTGGAGAAATTTACCACTAAACAGGTACCAACCTTAACGCAACCTGCTGTAACGTG 3837  
 Qy 3840 ----- 3839  
 Db 3838 AGATNACCGCTCAAAACAGTAGTATCTTAGTGGAAATGAGTCCAGCTGCTGCTGTAA 3897  
 Qy 3840 ----- 3839  
 Db 3898 CACTTACTGCAACCGAGGGGCTCTTGTCTGTAAGCAATATTTTCGGGCAACACCGTTACTG 3957  
 Qy 3840 ----- 3839  
 Db 3958 TTACTGCAATAGCGGTGCAATTAACCACTTTTGGCAGGCTCTACAATTAAGGAACCGAGA 4017  
 Qy 3840 -AGCAAGTATTACAAACCAACAGGTGATATACGCGGTACGATTTCCGGTAAACACGCTAA 3898  
 Db 4018 GTGTACCCTTCAAGTCAATCAGCGCATATCGCGGTACGATTTCTGCTGGCACACAGTAG 4077  
 Qy 3899 GTGTAGCGGCTGCTGATTTTAAACCACTAAATTCGGCTCAAAATTTGAAGCCAAATCGG 3958  
 Db 4078 AGGTAAAGCAACCGCAAGTTTAAACCACTCAATCAATTTCAAAATTTAAAGCAACACAG 4137  
 Qy 3959 GTGAGGCTATGTAACAAAGTGCACAGGTACAAATTTGGCGGTACAAATTTCCGGTAAACCG 4018  
 Db 4138 GCGAGGCTAACGTAAACAAAGTGCACAGGTACAAATTTGGTGTACGATTTTCCGGTAAACCG 4197

QY 4019 TAAATGTTTACGCAACGCTGGCGATTTACAGTTGGGAATGGCGCAGAAAATTAATGCCGA 4078  
Db 4198 TAAATGTTTACGCAACGCTGGCGATTTACAGTTGGGAATGGCGCAGAAAATTAATGCCGA 4257  
QY 4079 CAGAAGAGCTGCAACCTTAACCGCAACAGGAATACCTTGACTACTGAAAGCCGGTTCTA 4138  
Db 4258 CAGAAGAGCTGCAACCTTAACCTACATCATCGGGCAAAATTAACCTCCGAAGCTAGTTTCA 4317  
QY 4139 GCATCACTTCAACTAAGGGTCAGTAGACCTCTTGCTGCTCAGAATGGTAGCATCGCAGGAA 4198  
Db 4318 ACATTAATGCTGCTTAATGTCACATTAATTAACCTACAGCCACCTTAACCCAGCTGGCAGGCT 4258  
QY 4199 GCATTAATGCTGCTTAATGTCACATTAATTAACCTACAGCCACCTTAACCCAGCTGGCAGGCT 4258  
Db 4378 GTATTAATGCTGCTTAATGTCACATTAATTAACCTACAGCCACCTTAACCCAGCTGGCAGGCT 4437  
QY 4259 CGGATATTAAGCAACAGCGGACCTTGTTTATTAAACGCAAAAGATGCTAAAGCTAAATG 4318  
Db 4438 CAACATATTAAGCAACAGCGGACCTTGTTTATTAAACGCAAAAGATGCTAAAGCTAAATG 4497  
QY 4319 GTGATGCATCAGGTGATGATACAGAGTCAATGCAGTCAACGCAAGCGGCTCTGGTAGTG 4378  
Db 4498 GCGCAGCATTTGGTATACCAACAGGTGATTAATGCAACCAACGCAAAATGGCTCCGCGACGG 4557  
QY 4379 TGACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGT 4438  
Db 4558 TAATCGGCAACCTCAAGCAGTGAATATCACTGGGATTTAAACACAGTAAATGGT 4617  
QY 4439 TAAATATCAATTTTCAAGAGTGTGAACACACACACACACACACACACACACACACACAC 4498  
Db 4618 TAAATATCAATTTTCAAGAGTGTGAACACACACACACACACACACACACACACACACAC 4677  
QY 4499 TGAATATATCCAGCAGGTGTAGCAAGTGTAGAGAGTGTAGAGAGTGTAGAGAGTGTAGAG 4558  
Db 4678 TGAATATATCCAGCAGGTGTAGCAAGTGTAGAGAGTGTAGAGAGTGTAGAGAGTGTAGAG 4737  
QY 4559 TTGAAAAGTAAAGATTTATCTGATGAAGAAAGAAACACATTAAGTAAACTTGGTGTA 4618  
Db 4738 TTGAAAAGTAAAGATTTATCTGATGAAGAAAGAAACACATTAAGTAAACTTGGTGTA 4797  
QY 4619 GTGCTGTACGTTTGTGAGCAATATCAATTAACATTAACATTAACATTAACATTAACATTA 4678  
Db 4798 GTGCTGTACGTTTGTGAGCAATATCAATTAACATTAACATTAACATTAACATTAACATTA 4857  
QY 4679 CAACGACACCTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTGTCTCAAGTGA 4738  
Db 4858 CAACGACACCTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTGTCTCAAGTGA 4917  
QY 4739 ATGGCGACGATGTACCAATGTTGTCAGGATGGACACACACACACACACACACACACAC 4798  
Db 4918 ATGGCGACGATGTACCAATGTTGTCAGGATGGACACACACACACACACACACACACAC 4977  
QY 4799 AGGTAGATTTTCACTGCAATGAAGTCAATTTTTCGTTATTTTTCGTTATTTTTCGTTATTT 4858  
Db 4978 AGGTAGATTTTCACTGCAATGAAGTCAATTTTTCGTTATTTTTCGTTATTTTTCGTTATTT 5037  
QY 4859 AAGTTCAGTACGGGCTTTACCCATCTTGTAAAAAATTTACGGAGATACAAATAAAGTATTT 4918  
Db 5038 AAGTTCAGTACGGGCTTTACCCATCTTGTAAAAAATTTACGGAGATACAAATAAAGTATTT 5097  
QY 4919 TTAACAGGTTATTATTATG 4937  
Db 5098 TTAACAGGTTATTATTATG 5116

Search completed: March 22, 2003, 00:41:58  
Job time : 12519 secs

us-10-092-880-3.rng

Tue Mar 25 08:39:01 2003

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 15:43:31 ; Search time 909 Seconds  
(without alignments)  
12231.150 Million cell updates/sec

Title: US-10-092-880-3  
Perfect score: 4937  
Sequence: 1 taatatacagaataataa.....tttaacagggtattattatg 4937

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 437078

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :	N_Geneseq_101002.*
1:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4937	100.0	4937	18	Non-typeable Haemo
2	4935.4	100.0	4937	14	Sequence encoding
3	4924	99.7	4940	21	Haemophilus influe
4	4913	99.5	4937	15	Haemophilus high m
5	4828.2	97.8	9323	14	Gene cluster for h
6	4820.2	97.6	9323	18	Non-typeable Haemo
7	3263	66.1	3263	21	DNA encoding Haemo
8	2680.2	54.3	5116	14	Sequence encoding
9	2680.2	54.3	5116	15	Haemophilus high m

10	2680.2	54.3	5116	18	AAT90994	Non-typeable Haemo
11	2680.2	54.3	5116	21	AAAS2195	Haemophilus influe
12	2677	54.2	9220	14	AAQ49508	Gene cluster for h
13	2565.4	52.0	9171	18	AAT90996	Non-typeable Haemo
14	2245	45.5	4803	18	AAT90993	Non-typeable Haemo
15	2203.8	44.6	4702	14	AAQ49511	Partial sequence e
16	2202.2	44.6	4702	15	AAQ72296	Haemophilus high m
17	1974	40.0	4794	18	AAT90992	Non-typeable Haemo
18	1591.8	32.2	2874	21	AAAS2188	DNA encoding H. in
19	1591.8	32.2	2892	21	AAAS2187	Haemophilus influe
20	1500.8	30.4	3240	21	AAAS2185	Haemophilus influe
21	1497.6	30.3	2997	21	AAAS2191	Haemophilus influe
22	1484.4	30.1	3222	21	AAAS2186	DNA encoding H. in
23	1481.2	30.0	2979	21	AAAS2192	DNA encoding Haemo
24	1473.2	29.8	4287	15	AAQ72295	Haemophilus high m
25	1469.2	29.8	4286	14	AAQ49510	Partial sequence e
26	1414	28.6	2928	21	AAAS2177	Haemophilus influe
27	1397.6	28.3	2910	21	AAAS2178	DNA encoding H. in
28	1298.4	26.3	3018	21	AAAS2182	Haemophilus influe
29	1298.4	26.3	3036	21	AAAS2181	Haemophilus influe
30	1188.4	24.1	3033	21	AAAS2189	Haemophilus influe
31	1172	23.7	3015	21	AAAS2190	DNA encoding H. in
32	1125.4	22.8	3443	21	AAAS2196	DNA encoding Haemo
33	1081.2	21.9	3306	21	AAAS2183	Haemophilus influe
34	1068	21.6	3288	21	AAAS2184	DNA encoding Haemo
35	753.6	15.3	3543	21	AAAS2194	Haemophilus influe
36	753.6	15.3	3568	21	AAAS2193	Haemophilus influe
37	643.2	13.0	3667	21	AAAS2180	DNA encoding H. in
38	643.2	13.0	3685	21	AAAS2179	Haemophilus influe
39	631.4	12.8	3681	21	AAAS2175	Haemophilus influe
40	628.8	12.7	3663	21	AAAS2176	DNA encoding H. in
41	65	1.3	105	21	AAAS2165	Plasmid DS-1200-3
42	64.2	1.3	2270	23	ABL19882	Drosophila melanog
43	62.8	1.3	1830421	17	AAT42063	Haemophilus influe
44	61	1.2	105	21	AAAS2164	Plasmid DS-1200-3
45	60.8	1.2	112190	22	AH44801	Human GPCR protein

## ALIGNMENTS

RESULT 1	AAT90995	standard; DNA: 4937 BP.
ID	AAT90995	Non-typeable Haemophilus high mol.wt. surface protein hmw2 gene.
XX		Non-typeable Haemophilus; high molecular weight surface protein;
AC	AAT90995;	HMW2; hmw2 gene; immunogen; vaccine; otitis media; ss.
XX		Haemophilus influenzae strain 12.
DT	14-APR-1998	(first entry)
XX		Location/Qualifiers
XX	Key	341...345
FH	RBS	/tag= a
FT	repeat_unit	129...135
FT		/rpt_type= b
FT	repeat_unit	136...142
FT		/tag= c
FT	repeat_unit	143...149
FT		/rpt_type= d
FT	repeat_unit	150...156
FT		/tag= e
FT	repeat_unit	157...163
FT		/tag= f

```

FT      repeat_unit      /rpt_type= TANDEM
FT      164..170          /*tag= g
FT      repeat_unit      /rpt_type= TANDEM
FT      171..177          /*tag= h
FT      repeat_unit      /rpt_type= TANDEM
FT      178..184          /*tag= i
FT      repeat_unit      /rpt_type= TANDEM
FT      185..191          /*tag= j
FT      repeat_unit      /rpt_type= TANDEM
FT      192..198          /*tag= k
FT      repeat_unit      /rpt_type= TANDEM
FT      199..205          /*tag= l
FT      repeat_unit      /rpt_type= TANDEM
FT      206..212          /*tag= m
FT      repeat_unit      /rpt_type= TANDEM
FT      213..219          /*tag= n
FT      repeat_unit      /rpt_type= TANDEM
FT      220..226          /*tag= o
FT      repeat_unit      /rpt_type= TANDEM
FT      227..233          /*tag= p
FT      repeat_unit      /rpt_type= TANDEM
FT      234..240          /*tag= q
FT      CDS               /rpt_type= TANDEM
FT      352..478          /*tag= i
FT      /transl_except= (pos:451..453, aa:Ser)
FT      /transl_except= (pos:454..456, aa:Glu)
FT      /transl_except= (pos:457..459, aa:Lys)
FT      /transl_except= (pos:460..462, aa:Pro)
FT      /transl_except= (pos:463..465, aa:Ala)
FT      /transl_except= (pos:466..468, aa:Arg)
FT      /transl_except= (pos:469..471, aa:Met)
FT      /transl_except= (pos:472..474, aa:Lys)
FT      /transl_except= (pos:475..477, aa:Val)
FT      /transl_except= (pos:478..480, aa:Arg)
FT      /transl_except= (pos:1621..1623, aa:Asp)
FT      /transl_except= (pos:1627..1629, aa:Phe)
FT      /transl_except= (pos:1633..1635, aa:Lys)
FT      /transl_except= (pos:1636..1638, aa:Asp)
FT      /transl_except= (pos:1651..1653, aa:Asp)
FT      /transl_except= (pos:1654..1656, aa:Ala)
FT      /transl_except= (pos:1675..1677, aa:Phe)
FT      /transl_except= (pos:1681..1683, aa:Asn)
FT      /transl_except= (pos:1687..1689, aa:Ser)
FT      /transl_except= (pos:1693..1695, aa:Asn)
FT      /transl_except= (pos:1711..1713, aa:Phe)
PN      W09736914-AL.
XX
PD      09-OCT-1997.
XX
PF      01-APR-1997; 97WO-US04707.
XX
PR      01-APR-1996; 96US-0617697.
XX
PA      (BARE/) BARENKAMP S J.
XX
PI      Barenkamp SJ;
XX
DR      WPI; 1997-503038/46.
DR      P-PSDB; AAW30294.
XX

```

```

PT      High molecular weight proteins of non-typeable Haemophilus
PT      Influenzae - useful for vaccine production
XX
PS      Claim 7; Page 71-73; 183pp; English.
XX
CC      This nucleic acid comprises the hmw2 gene of non-typeable
CC      Haemophilus influenzae strain 12 that encodes high molecular
CC      weight surface protein HMW2 (see AAW30294). A phage genomic
CC      library of strain 12 was screened for clones expressing high
CC      mol.wt. proteins using a high titre antiserum against HMWs.
CC      Strongly reactive clones were subcloned into p7 expression
CC      plasmid; all expressed either 120 kDa HMW2 or 125 kDa HMW1 (see
CC      AAW30293). The expressed proteins are truncated, starting at
CC      residue 442 of both full-length HMW1 and HMW2 gene products.
CC      Correct processing requires the products of additional
CC      downstream genes (see AAT90996 and AAT90997). Nucleotide sequences
CC      (see AAT90992-93) encoding 2 HMW proteins of non-typeable H.
CC      influenzae strain 5 (see AAW30291-92) have also been identified.
CC      The HMW proteins, conjugates and peptides can be used in,
CC      vaccines, as immunogens for preparation of antibodies and as
CC      antigens for detection of these antibodies. The nucleic acid
CC      sequences can be used as to prepare recombinant proteins and as
CC      probes for detection of related genes.
XX
SQ      Sequence 4937 BP; 1729 A; 948 C; 1009 G; 1251 T; 0 other;

Query Match      100.0%; Score 4937; DB 18; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TAAATATACAAGATAATAAATAAATAAATCAAGATTTTGTGTATGACAAACACAAATTACAA 60
Db      1 TAAATATACAAGATAATAAATAAATAAATCAAGATTTTGTGTATGACAAACACAAATTACAA 60
Qy      61 CACCTTTTTCAGCTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCATATAA 120
Db      61 CACCTTTTTCAGCTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCATATAA 120
Qy      121 AATGGTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 180
Db      121 AATGGTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 180
Qy      181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Db      181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Qy      241 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db      241 CACATGAATATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Qy      301 GAACGCAATATGATAAGTAAATTTAATTTGTTCAACTAACCTTAGGAGAAATATGAACAAG 360
Db      301 GAACGCAATATGATAAGTAAATTTAATTTGTTCAACTAACCTTAGGAGAAATATGAACAAG 360
Qy      361 ATATATCGTCTCAAAATTCAGCAACGCCCTGAATGCTTTGGTGGTGTGTGTGTGAATGGCA 420
Db      361 ATATATCGTCTCAAAATTCAGCAACGCCCTGAATGCTTTGGTGGTGTGTGTGTGAATGGCA 420
Qy      421 CGGGGTTGTGACCAATTCACAGAAAAGGCTTCCCGCTATGTTTACTATCTTTTAGGTGTAC 480
Db      421 CGGGGTTGTGACCAATTCACAGAAAAGGCTTCCCGCTATGTTTACTATCTTTTAGGTGTAC 480
Qy      481 CACTTAGCGTTAAAGCCACTTTCCCGCTATGTTTACTATCTTTTAGGTGTATTCCTCA 540
Db      481 CACTTAGCGTTAAAGCCACTTTCCCGCTATGTTTACTATCTTTTAGGTGTATTCCTCA 540
Qy      541 CAATCTGTTTATAGCAAGGGCTTTACAGGAATGGATGTAGTACACGCGACGCCACTATG 600
Db      541 CAATCTGTTTATAGCAAGGGCTTTACAGGAATGGATGTAGTACACGCGACGCCACTATG 600
Qy      601 CAAAGTAGATGTTAATAAACCATTATCCGCAACAGCTGTTGACGCTATCATTAATTTGGAAA 660
Db      601 CAAAGTAGATGTTAATAAACCATTATCCGCAACAGCTGTTGACGCTATCATTAATTTGGAAA 660

```

us-10-092-880-3.rng

Tue Mar 25 08:39:01 2003

QY	661	CAATTTAACTGACCAAAATGAATGGTCAGTGTCTTACAAAGAAACAACTCGGC	720
Db	661	CAATTTAACTGACCAAAATGAATGGTCAGTGTCTTACAAAGAAACAACTCGGC	720
QY	721	GTATTTCAACCGTGTAGATCTTAACCAATCTCCCAATTTAAAGAGGATTTTAGATTTAAAC	780
Db	721	GTATTTCAACCGTGTAGATCTTAACCAATCTCCCAATTTAAAGAGGATTTTAGATTTAAAC	780
QY	781	GGCAAGTCTTTTAAATCAACCAANTGGTATCAATAGTAAAGAGCAATTTATTAAAC	840
Db	781	GGCAAGTCTTTTAAATCAACCAANTGGTATCAATAGTAAAGAGCAATTTATTAAAC	840
QY	841	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTTAACGAAACATCAAGCGCGTAAT	900
Db	841	ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTTAACGAAACATCAAGCGCGTAAT	900
QY	901	TTCACTTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTAATTCACGGTTAAAT	960
Db	901	TTCACTTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTAATTCACGGTTAAAT	960
QY	961	ACTGTGGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGAAACGAGGGTGT	1020
Db	961	ACTGTGGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGAAACGAGGGTGT	1020
QY	1021	ATTAGCGTAATGGTGGCAGCATTTCTTACTCGCAGGCAAAAAATCACCATCAGCGAT	1080
Db	1021	ATTAGCGTAATGGTGGCAGCATTTCTTACTCGCAGGCAAAAAATCACCATCAGCGAT	1080
QY	1081	ATAATAACCCCAACCATTTACTTACAGATTTGCCGCCCTGAAAAATGAAGCGGTCAATCTG	1140
Db	1081	ATAATAACCCCAACCATTTACTTACAGATTTGCCGCCCTGAAAAATGAAGCGGTCAATCTG	1140
QY	1141	GGCGATTTTTCGCAAGGCGTAACTTAATGTCGCTGCTGCCACTATTTCGAAACCAA	1200
Db	1141	GGCGATTTTTCGCAAGGCGTAACTTAATGTCGCTGCTGCCACTATTTCGAAACCAA	1200
QY	1201	GGTAACTTTCTGCTGATCTGTAAAGCAAGATAAAGCGCAATTTGTTCTTCGCGC	1260
Db	1201	GGTAACTTTCTGCTGATCTGTAAAGCAAGATAAAGCGCAATTTGTTCTTCGCGC	1260
QY	1261	AAAGAGGTGAAGCGGAAATTTGGCGGTGTAAATTTCCGCTCAAAATCAGCAAGTAAAGC	1320
Db	1261	AAAGAGGTGAAGCGGAAATTTGGCGGTGTAAATTTCCGCTCAAAATCAGCAAGTAAAGC	1320
QY	1321	GGCAAGCTGATGATTTACAGCGGATAAAGTCAATTTAAACACAGGTGAGTTATCGACCTT	1380
Db	1321	GGCAAGCTGATGATTTACAGCGGATAAAGTCAATTTAAACACAGGTGAGTTATCGACCTT	1380
QY	1381	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGGTAAAC	1440
Db	1381	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCAAGGTAAAC	1440
QY	1441	GGCATCAATTTAGCAAGAAACCTTTTAAAGAAAGCTTCAACCATCAATGTATCAGGC	1500
Db	1441	GGCATCAATTTAGCAAGAAACCTTTTAAAGAAAGCTTCAACCATCAATGTATCAGGC	1500
QY	1501	AAAGAAAGGGGAGCGCTATTTGTTGGGCGGATATTGGCTTAATTAAGCGCAATATT	1560
Db	1501	AAAGAAAGGGGAGCGCTATTTGTTGGGCGGATATTGGCTTAATTAAGCGCAATATT	1560
QY	1561	AACGCTCAAGGTAGTGGTGATTCGCTAAACCGGTGTTTGTGAGACATTCGGGGCAT	1620
Db	1561	AACGCTCAAGGTAGTGGTGATTCGCTAAACCGGTGTTTGTGAGACATTCGGGGCAT	1620
QY	1621	TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGTGGTGTGACGCTCAT	1680
Db	1621	TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAGAGTGGTGTGACGCTCAT	1680
QY	1681	GATGTAAACAATTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATAAATGATGAATTC	1740
Db	1681	GATGTAAACAATTGAAGCGGAGACCCCTTTCGCAATTAATACCGGTATAAATGATGAATTC	1740
QY	1741	CCAACGCGACCGTGAAGCAAGCGACCTTAAAAAAAATAGCGAACTCAAAACACGCTA	1800
Db	1741	CCAACGCGACCGTGAAGCAAGCGACCTTAAAAAAAATAGCGAACTCAAAACACGCTA	1800
QY	1801	ACCAATCAACATTAATTTCAAAATTTATCTGAAAAACGCTTGGACAAATGAATATAAGCGCATCA	1860
Db	1801	ACCAATCAACATTAATTTCAAAATTTATCTGAAAAACGCTTGGACAAATGAATATAAGCGCATCA	1860
QY	1861	AGAAAACCTTACCGTTAATAGCTCAATCAACATCGAAGCAACTCCCACTTTAAATTTCCAT	1920
Db	1861	AGAAAACCTTACCGTTAATAGCTCAATCAACATCGAAGCAACTCCCACTTTAAATTTCCAT	1920
QY	1921	AGTAAAGGTTCAGCGTGGCGGAGCGTTCAGATTGATGAGATATTACTTCTTAAAGCGGGA	1980
Db	1921	AGTAAAGGTTCAGCGTGGCGGAGCGTTCAGATTGATGAGATATTACTTCTTAAAGCGGGA	1980
QY	1981	AATTTAACCATTTAATTTCTGGCGGATGGGTGATGTTTCAAAAAATATTACGCTTTGATCAG	2040
Db	1981	AATTTAACCATTTAATTTCTGGCGGATGGGTGATGTTTCAAAAAATATTACGCTTTGATCAG	2040
QY	2041	GGTTTTTTTAAATTAATTTACCGCGCTTCCGCTAGCTTTTGAAGGTGGAAATTAACAAAGCAGC	2100
Db	2041	GGTTTTTTTAAATTAATTTACCGCGCTTCCGCTAGCTTTTGAAGGTGGAAATTAACAAAGCAGC	2100
QY	2101	GACGCGCAAAATGCTAAAAATTTGCGCCAGGCGACTGTAAACCATTTACAGGAGAGGAAAA	2160
Db	2101	GACGCGCAAAATGCTAAAAATTTGCGCCAGGCGACTGTAAACCATTTACAGGAGAGGAAAA	2160
QY	2161	GATTTACAGGCTTAAACAGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA	2220
Db	2161	GATTTACAGGCTTAAACAGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA	2220
QY	2221	TCAGTGAATTAATTTAACCCACAACTTAGTGGCACAATTAACATATCTGGGAATATAACA	2280
Db	2221	TCAGTGAATTAATTTAACCCACAACTTAGTGGCACAATTAACATATCTGGGAATATAACA	2280
QY	2281	ATTAACCAAACTAGGAGAAAGACACCTGATTTGGCAACACAGCCATGATTCGCACTGG	2340
Db	2281	ATTAACCAAACTAGGAGAAAGACACCTGATTTGGCAACACAGCCATGATTCGCACTGG	2340
QY	2341	AACGTCAGTGTCTTAATCTAGAGACAGCGGCAAAATTTTACCTTTTAAATATCATTTCA	2400
Db	2341	AACGTCAGTGTCTTAATCTAGAGACAGCGGCAAAATTTTACCTTTTAAATATCATTTCA	2400
QY	2401	AGCAATAGCAAAAGGCTTAAACACACAGTATAGAAGCTCTGCAGGGGTGAATTTTAAAGGC	2460
Db	2401	AGCAATAGCAAAAGGCTTAAACACACAGTATAGAAGCTCTGCAGGGGTGAATTTTAAAGGC	2460
QY	2461	GTAAATGGCAACATGTCATTTCAATCTCAAGAGGAGCGGAAAGTTAATTTCAATTTAA	2520
Db	2461	GTAAATGGCAACATGTCATTTCAATCTCAAGAGGAGCGGAAAGTTAATTTCAATTTAA	2520
QY	2521	CCAAACGAGAACATCAACACAGCAAACTTTTACCAATTCGGTTTTTGTAGCAATATCACA	2580
Db	2521	CCAAACGAGAACATCAACACAGCAAACTTTTACCAATTCGGTTTTTGTAGCAATATCACA	2580
QY	2581	GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCAACCACTTCTGGCAGAGGGCT	2640
Db	2581	GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCAACCACTTCTGGCAGAGGGCT	2640
QY	2641	GAGTTAAAAATGAGTGAAATTAATCTTCAACGGCGCTAATTTTACCTTAAATTTCCCAT	2700
Db	2641	GAGTTAAAAATGAGTGAAATTAATCTTCAACGGCGCTAATTTTACCTTAAATTTCCCAT	2700
QY	2701	GTTTCGGGGGATGACGCTTTTAAATCAACAGACTTAACCATTAATTCGAACCAATTTCA	2760
Db	2701	GTTTCGGGGGATGACGCTTTTAAATCAACAGACTTAACCATTAATTCGAACCAATTTCA	2760
QY	2761	AATTTACGCTCAGACAGAGGAAAGATTTTATGACGGGTACGAGCAATGCCATC	2820
Db	2761	AATTTACGCTCAGACAGAGGAAAGATTTTATGACGGGTACGAGCAATGCCATC	2820
QY	2821	AATTTACCTTACACATATCCATTTCTGGCGGTAAATGTCCACCTTGGTGGGCAAAACTCA	2880
Db	2821	AATTTACCTTACACATATCCATTTCTGGCGGTAAATGTCCACCTTGGTGGGCAAAACTCA	2880



Db 2821 AATTCACCTCAACATATCCATTCTGGCGGTAAATGTCCACCTTGGTGGACAAACTCA 2880  
QY 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAGCAGCAAAATGTTACGCTAGAAGCC 2940  
Db 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAGCAGCAAAATGTTACGCTAGAAGCC 2940  
QY 2941 AATAACGCCCTTAATCAGCAAAACATAAGGGATAGAGTATATAAATTTGGCAGCTTGCTC 3000  
Db 2941 AATAACGCCCTTAATCAGCAAAACATAAGGGATAGAGTATATAAATTTGGCAGCTTGCTC 3000  
QY 3001 GTTAATGGAGTTTAAGTTTAACCTGGCGAAATGCAGATATAAGGCAATCTCACTATT 3060  
Db 3001 GTTAATGGAGTTTAAGTTTAACCTGGCGAAATGCAGATATAAGGCAATCTCACTATT 3060  
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCAACCGGCAATTTT 3120  
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCCTAAATATCAACCGGCAATTTT 3120  
QY 3121 ACCAATAATGGCACTGCGGAAATTAATATACCAAGAGTGTAAACCTTGGCAATGTT 3180  
Db 3121 ACCAATAATGGCACTGCGGAAATTAATATACCAAGAGTGTAAACCTTGGCAATGTT 3180  
QY 3181 ACCAATGATGTTTAAACATATACCACTACCGCTAAAGCAACCAAGAGCATCATC 3240  
Db 3181 ACCAATGATGTTTAAACATATACCACTACCGCTAAAGCAACCAAGAGCATCATC 3240  
QY 3241 GCGGAGATATAATCAACAAAGAAAGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300  
Db 3241 GCGGAGATATAATCAACAAAGAAAGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300  
QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAAAGGCAACCTCACGATTTCTCCGAT 3360  
Db 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAAAGGCAACCTCACGATTTCTCCGAT 3360  
QY 3361 AAAATTAATATACCAAAACAGATACCAATCAAAAGGGTATTGATGGAGAGGACTCTAGT 3420  
Db 3361 AAAATTAATATACCAAAACAGATACCAATCAAAAGGGTATTGATGGAGAGGACTCTAGT 3420  
QY 3421 TCAGATGCCACAGTAATGCCAACCTAATATTAAGCAAGAAATGAAATTTGACAGAA 3480  
Db 3421 TCAGATGCCACAGTAATGCCAACCTAATATTAAGCAAGAAATGAAATTTGACAGAA 3480  
QY 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCAAAAGATGTTAGAGATTTA 3540  
Db 3481 GACCTAAGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCAAAAGATGTTAGAGATTTA 3540  
QY 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCCGAAGCCAAACAGTAACCTTTAAC 3600  
Db 3541 ACTATTGGCAACAGTAATGACGGTAAACAGCGGTGCCGAAGCCAAACAGTAACCTTTAAC 3600  
QY 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTACAAATGACACATGACACTTAAATAGCAAGTG 3660  
Db 3601 AATGTTAAAGATTCAAAATCTCTGCTGACGGTACAAATGACACATGACACTTAAATAGCAAGTG 3660  
QY 3661 AAAACATCTAGCAATAGGCGAGCTGAAAGCAATAGCCAGCAAGATACCGGCTTAAC 3720  
Db 3661 AAAACATCTAGCAATAGGCGAGCTGAAAGCAATAGCCAGCAAGATACCGGCTTAAC 3720  
QY 3721 ATTACTGCAAAAATGATAGAGTAAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780  
Db 3721 ATTACTGCAAAAATGATAGAGTAAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780  
QY 3781 ACCGCGTGGAAAGGTTACCAACAGCAGGCTCGACCATTAACCAACAAATGGCAAA 3840  
Db 3781 ACCGCGTGGAAAGGTTACCAACAGCAGGCTCGACCATTAACCAACAAATGGCAAA 3840  
QY 3841 GCAAGTATTACACCAACAGGTTATACCGGATGATATCGCGTAAACAGGTAAGT 3900  
Db 3841 GCAAGTATTACACCAACAGGTTATACCGGATGATATCGCGTAAACAGGTAAGT 3900  
QY 3901 GTTAGCGGACTGGTATTAAACCACTAATCGGCTCAAAATTTGAAGCGAAATCGGGT 3960  
Db 3901 GTTAGCGGACTGGTATTAAACCACTAATCGGCTCAAAATTTGAAGCGAAATCGGGT 3960

Db 3901 GTTAGCGGACTGGTATTAAACCACTAATCGGCTCAAAATTTGAAGCGAAATCGGGT 3960  
QY 3961 GAGGCTAATGTAAACAAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAATACGGTA 4020  
Db 3961 GAGGCTAATGTAAACAAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGGTAATACGGTA 4020  
QY 4021 AATGTTACGCCAACCGTGGCGATTTAAACAGTTTGGGAATGGCGCAGAAATTAATGGGACA 4080  
Db 4021 AATGTTACGCCAACCGTGGCGATTTAAACAGTTTGGGAATGGCGCAGAAATTAATGGGACA 4080  
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGCTTCTAGC 4140  
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGCTTCTAGC 4140  
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTCTGGCTCACAATGGTAGCATCGCAGAAAGC 4200  
Db 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTTCTGGCTCACAATGGTAGCATCGCAGAAAGC 4200  
QY 4201 ATTAATGCTGCTAATGTGACATTAATACTACAGGCACCTTAACCCCTGGCAGGCTCG 4260  
Db 4201 ATTAATGCTGCTAATGTGACATTAATACTACAGGCACCTTAACCCCTGGCAGGCTCG 4260  
QY 4261 GATATTAAGCAACCGAGGGCACTTGGTATTAAACGCAAAAGATGCTAAGCTAAATGTT 4320  
Db 4261 GATATTAAGCAACCGAGGGCACTTGGTATTAAACGCAAAAGATGCTAAGCTAAATGTT 4320  
QY 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380  
Db 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380  
QY 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATAGGTTA 4440  
Db 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAATAGGTTA 4440  
QY 4441 AATATCAATTCGAAAGATGGTAGAAGTGTAGCAAGTGAAGAGGCAAGAAATGAGGTG 4500  
Db 4441 AATATCAATTCGAAAGATGGTAGAAGTGTAGCAAGTGAAGAGGCAAGAAATGAGGTG 4500  
QY 4501 AAATATATCCAGCGAGTGTAGCAAGTGTAGCAAGTGAAGAGGCAAGAAATGAGGTG 4560  
Db 4501 AAATATATCCAGCGAGTGTAGCAAGTGTAGCAAGTGAAGAGGCAAGAAATGAGGTG 4560  
QY 4561 GAAAGCTAAAGATTTATCTGATGAGAAAGAGAAACATTAAGCTTGGTGAAGT 4620  
Db 4561 GAAAGCTAAAGATTTATCTGATGAGAAAGAGAAACATTAAGCTTGGTGAAGT 4620  
QY 4621 GCTGTACGTTTCTGAGCCAAATATAACATTAACAGTCAATACAGCAAAATGAATTTACA 4680  
Db 4621 GCTGTACGTTTCTGAGCCAAATATAACATTAACAGTCAATACAGCAAAATGAATTTACA 4680  
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 4740  
Db 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAGGTAAGCGGTGTTCTCAAGTGGTAAT 4740  
QY 4741 GGCGCACAGTATGTACCAATGTCTGACAGTGAACAGCCGTAGTCAAGTAAATGACAAAG 4800  
Db 4741 GGCGCACAGTATGTACCAATGTCTGACAGTGAACAGCCGTAGTCAAGTAAATGACAAAG 4800  
QY 4801 GTAGATTTCTCATCTGCAATGAAGTCAATTTATTTTCGTTATTTTACGTGGGTAAA 4860  
Db 4801 GTAGATTTCTCATCTGCAATGAAGTCAATTTATTTTCGTTATTTTACGTGGGTAAA 4860  
QY 4861 GTTACAGTACGGCTTTTACCCTCTCTGTAATAAATTTACGGAGAATACAAATGATTTTT 4920  
Db 4861 GTTACAGTACGGCTTTTACCCTCTCTGTAATAAATTTACGGAGAATACAAATGATTTTT 4920  
QY 4921 AACAGTTTATTATTATG 4937  
Db 4921 AACAGTTTATTATTATG 4937

us-10-092-880-3.rng

Tue Mar 25 08:39:01 2003

```

ID AAQ49507 standard; DNA; 4937 BP.
XX AC AAQ49507;
XX DT 26-APR-1994 (first entry)
XX DE Sequence encoding high molecular weight protein 2 (HMW2).
XX KW HMW; high molecular weight protein; virus; vaccine; influenza;
XX KW epitope; immunity; haemophilus influenzae; ss.
XX OS Haemophilus influenzae.
XX Key Location/Qualifiers
XX FH 352..4785
XX CD /product= a
XX FT /product= High molecular weight protein 2.
XX FT misc_difference 451..453
XX FT /tag= b
XX FT /transl_except= TTC encodes Serine.
XX FT misc_difference 454..456
XX FT /tag= c
XX FT /transl_except= CGC encodes Glutamic acid.
XX FT misc_difference 457..459
XX FT /tag= d
XX FT /transl_except= TAT encodes Lysine.
XX FT misc_difference 460..462
XX FT /tag= e
XX FT /transl_except= GTT encodes Proline.
XX FT misc_difference 463..465
XX FT /tag= f
XX FT /transl_except= ACT encodes Alanine.
XX FT misc_difference 466..468
XX FT /tag= g
XX FT /transl_except= ATC encodes Arginine.
XX FT misc_difference 469..471
XX FT /tag= h
XX FT /transl_except= TTT encodes Methionine.
XX FT misc_difference 472..474
XX FT /tag= i
XX FT /transl_except= AGG encodes Lysine.
XX FT misc_difference 475..477
XX FT /tag= j
XX FT /transl_except= TGT encodes Valine.
XX FT misc_difference 478..480
XX FT /tag= k
XX FT /transl_except= AAC encodes Arginine.
XX FT misc_difference 1621..1623
XX FT /tag= l
XX FT /transl_except= TAT encodes Aspartic acid.
XX FT misc_difference 1627..1629
XX FT /tag= m
XX FT /transl_except= TCC encodes Phenylalanine.
XX FT misc_difference 1633..1635
XX FT /tag= n
XX FT /transl_except= GAC encodes Lysine.
XX FT misc_difference 1636..1638
XX FT /tag= o
XX FT /transl_except= AGC encodes Aspartic acid.
XX FT misc_difference 1651..1653
XX FT /tag= p
XX FT /transl_except= AAA encodes Aspartic acid.
XX FT misc_difference 1654..1656
XX FT /tag= q
XX FT /transl_except= ACA encodes Alanine.
XX FT misc_difference 1675..1677
XX FT /tag= r
XX FT /transl_except= CCT encodes phenylalanine.
XX FT misc_difference 1681..1683
XX FT /tag= s
XX FT /transl_except= GAT encodes Asparagine.
XX FT misc_difference 1687..1689
XX FT /tag= t

```

```

FT misc_difference 1693..1695
FT /transl_except= ACA encodes Serine.
FT /tag= u
FT /transl_except= GAA encodes Asparagine.
FT misc_difference 3913..3915
FT /tag= v
FT /transl_except= GGT encodes Valine.
XX WO9319090-A.
XX 30-SEP-1993.
XX 16-MAR-1993; 93WO-US02166.
XX 16-MAR-1992; 92GB-0005704.
XX (BARE/) BARENKAMP S J.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX Barenkamp SJ;
XX WPI: 1993-320683/40.
XX P-PSDB; AAR41724.
XX High molecular weight surface proteins - of non-typeable
XX haemophilus which exhibit immunogenic properties
XX Claim 4; Figure 3; 100pp; English.
XX The isolation and purification of the high molecular weight protein
XX enables the identification of the major protective epitopes of the
XX protein by conventional epitope mapping. These epitopes can then be
XX synthesised using standard techniques and incorporated into fully
XX synthetic or recombinant vaccines.
XX Sequence 4937 BP; 1729 A; 947 C; 1010 G; 1251 T; 0 other;
XX SQ Query Match 100.0%; Score 4935.4; DB 14; Length 4937;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 4936; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGCAACAAACAAATACAA 60
DB 1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGCAACAAACAAATACAA 60
QY 61 CACCTTTTTCAGCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCATATA 120
DB 61 CACCTTTTTCAGCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCATATA 120
QY 121 AATGATATAATCTTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAAT 180
DB 121 AATGATATAATCTTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAAT 180
QY 181 CTTTCATCTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTT 240
DB 181 CTTTCATCTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTTTTCATCTTT 240
QY 241 CACATGAATATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
DB 241 CACATGAATATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 301 GAACGCAATATGAATGAATTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCT 360
DB 301 GAACGCAATATGAATGAATTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCT 360
QY 361 ATATATCGTCTCAAAATTCAGCAACCCCTGAATGCTTGTCTGCTCAATTTGCA 420
DB 361 ATATATCGTCTCAAAATTCAGCAACCCCTGAATGCTTGTCTGCTCAATTTGCA 420
QY 421 CGGGGTTGTGACCATTTCCACAGAAAAGGCTCCCGTATGTTACTATCTTTAGTCTAAC 480
DB 421 CGGGGTTGTGACCATTTCCACAGAAAAGGCTCCCGTATGTTACTATCTTTAGTCTAAC 480

```

QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540  
Db 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540  
QY 541 CAATCTGTTTTAGCAAGCGGCTTACAAAGATGGATGTAGTACACGGCACACCCACTATG 600  
Db 541 CAATCTGTTTTAGCAAGCGGCTTACAAAGATGGATGTAGTACACGGCACACCCACTATG 600  
QY 601 CAAGTAGATGGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGGAA 660  
Db 601 CAAGTAGATGGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGGAA 660  
QY 661 CAATTTAAACATCGACCAAAATCAAAATGTCAGTTTTTACAAGAAAACAACAATCCGCC 720  
Db 661 CAATTTAAACATCGACCAAAATCAAAATGTCAGTTTTTACAAGAAAACAACAATCCGCC 720  
QY 721 GTATTCACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTAGATTCTAAC 780  
Db 721 GTATTCACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTAGATTCTAAC 780  
QY 781 GGACAAGTCTTTTTAATCAACCCAAATGGTATCAAAATAGGTAAAGACGCAATTAATAAC 840  
Db 781 GGACAAGTCTTTTTAATCAACCCAAATGGTATCAAAATAGGTAAAGACGCAATTAATAAC 840  
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900  
Db 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900  
QY 901 TTCACCTTCGAGCAACCAAGATAAGCGCTCGCTGAAATGTGTAATCAGCGTTTAAT 960  
Db 901 TTCACCTTCGAGCAACCAAGATAAGCGCTCGCTGAAATGTGTAATCAGCGTTTAAT 960  
QY 961 ACTGTCGGTAAAGACGCGAGTGAATCTTATTTGGTGGCAAGTGAAGAACGAGGCTG 1020  
Db 961 ACTGTCGGTAAAGACGCGAGTGAATCTTATTTGGTGGCAAGTGAAGAACGAGGCTG 1020  
QY 1021 ATTAGCTGAATGGTGGCAGCATTTCTTACTGCGAGGCAAAAAATCACCATCAGCGAT 1080  
Db 1021 ATTAGCTGAATGGTGGCAGCATTTCTTACTGCGAGGCAAAAAATCACCATCAGCGAT 1080  
QY 1081 ATATAAACCACCACTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140  
Db 1081 ATATAAACCACCACTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGTCAATCTG 1140  
QY 1141 GCGGATATTTTGGCAAGGCGGTAAATTAATGTCCGTGCTGCACTATTCGAAACCAA 1200  
Db 1141 GCGGATATTTTGGCAAGGCGGTAAATTAATGTCCGTGCTGCACTATTCGAAACCAA 1200  
QY 1201 GGTAACTTTCTGCTGATTTGCGGTTAAATTTCCGCTCAAAATCAGCAAGTAAAGGC 1260  
Db 1201 GGTAACTTTCTGCTGATTTGCGGTTAAATTTCCGCTCAAAATCAGCAAGTAAAGGC 1260  
QY 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTAAATTTCCGCTCAAAATCAGCAAGTAAAGGC 1320  
Db 1261 AAAGAGGTGAAGCGGAAATTTGGCGGTAAATTTCCGCTCAAAATCAGCAAGTAAAGGC 1320  
QY 1321 GGCAAGCTGATGATTTACAGCGGATTAAGTACATTAATAACAGGTGCGATTCGACCTT 1380  
Db 1321 GGCAAGCTGATGATTTACAGCGGATTAAGTACATTAATAACAGGTGCGATTCGACCTT 1380  
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCGAAGTAAAAAC 1440  
Db 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCGAAGTAAAAAC 1440  
QY 1441 GGCATTCATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500  
Db 1441 GGCATTCATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500  
QY 1501 AAAGAAAAGCGGCGCTATTTGTTGGGCGATATTTGCGTTAATTCACGCGCATATT 1560  
Db 1501 AAAGAAAAGCGGCGCTATTTGTTGGGCGATATTTGCGTTAATTCACGCGCATATT 1560  
QY 1561 AACGCTCAAGGTAGTGGTATTCGCTAAACCGGTGGTTTTGTGGAGACATCGGGCAT 1620

Db 1561 AACGCTCAAGGTAGTGGTATATCGCTAAACCGGTGGTTTTGTGGAGACATCGGGCAT 1620  
QY 1621 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTGTAGACCCGTAT 1680  
Db 1621 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTGTAGACCCGTAT 1680  
QY 1681 GATGTAACAATTTGAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGATTC 1740  
Db 1681 GATGTAACAATTTGAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGATTC 1740  
QY 1741 CCAACAGGACCCGTTGAAGCAAGCGACCCCTTAAACAAAAGAGTGGTGTAGACCCGTAT 1800  
Db 1741 CCAACAGGACCCGTTGAAGCAAGCGACCCCTTAAACAAAAGAGTGGTGTAGACCCGTAT 1800  
QY 1801 ACCAATACAACTATTTCAAAATTTATCTGAAACAGGCTGGCAATGAATATAACGCGATCA 1860  
Db 1801 ACCAATACAACTATTTCAAAATTTATCTGAAACAGGCTGGCAATGAATATAACGCGATCA 1860  
QY 1861 AGAAACTTACCGTTAATAGCTCAATCAATCGGAAGCAACTCCCACTTAATTCCTCAT 1920  
Db 1861 AGAAACTTACCGTTAATAGCTCAATCAATCGGAAGCAACTCCCACTTAATTCCTCAT 1920  
QY 1921 AGTAAAGGTCAGCGTGGCGGAGCGCTTCAATCAATCGGAAGCAACTCCCACTTAATTCCTCAT 1980  
Db 1921 AGTAAAGGTCAGCGTGGCGGAGCGCTTCAATCAATCGGAAGCAACTCCCACTTAATTCCTCAT 1980  
QY 1981 AATTTAAACATTTATCTGGCGGATGGTGTGATGTTCAATAAATATTTACGCTTGTATCAG 2040  
Db 1981 AATTTAAACATTTATCTGGCGGATGGTGTGATGTTCAATAAATATTTACGCTTGTATCAG 2040  
QY 2041 GGTTTTTTAAATATTTACCGCGCTTCCGTAGCTTTTGAAGGTGGAAATAACAAAGCAGC 2100  
Db 2041 GGTTTTTTAAATATTTACCGCGCTTCCGTAGCTTTTGAAGGTGGAAATAACAAAGCAGC 2100  
QY 2101 GACGCGCAATGCTAAATTTGTCGCCAGGCGCTGTAACCATACAGGAGAGGAGAA 2160  
Db 2101 GACGCGCAATGCTAAATTTGTCGCCAGGCGCTGTAACCATACAGGAGAGGAGAA 2160  
QY 2161 GATTTCAAGGCTACACAGCTATCTTTAAAGGCAAGGCTGAAAGTCTGAAATATCAATTC 2220  
Db 2161 GATTTCAAGGCTACACAGCTATCTTTAAAGGCAAGGCTGAAAGTCTGAAATATCAATTC 2220  
QY 2221 TCAGTGAATATTTAAACCCACAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280  
Db 2221 TCAGTGAATATTTAAACCCACAATCTTAGTGGCACAATTAACATATCTGGGAATATAACA 2280  
QY 2281 ATTAACCAAACTAGGAGAAAGACACCTCGTATTGGCAACACCGCATGATTCGCACTGG 2340  
Db 2281 ATTAACCAAACTAGGAGAAAGACACCTCGTATTGGCAACACCGCATGATTCGCACTGG 2340  
QY 2341 AACGTCAGTCTCTTAATCTAGAGCAGCGCAATTTTACCTTTATTAATATCAATTTCA 2400  
Db 2341 AACGTCAGTCTCTTAATCTAGAGCAGCGCAATTTTACCTTTATTAATATCAATTTCA 2400  
QY 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGCAAGGCTGAAATTTAACGCG 2460  
Db 2401 AGCAATAGCAAGGCTTAAACACACAGTATAGAAGCTCTGCAAGGCTGAAATTTAACGCG 2460  
QY 2461 GTAAATGGCAACATGTCATTTCAATCTCAAGAGGAGGGAAGTAAATTTCAATTTAA 2520  
Db 2461 GTAAATGGCAACATGTCATTTCAATCTCAAGAGGAGGGAAGTAAATTTCAATTTAA 2520  
QY 2521 CCAACAGGAGACATGAACACACAGCAAACTTTACCAATTCGTTTTAGCCAAATATCACA 2580  
Db 2521 CCAACAGGAGACATGAACACACAGCAAACTTTACCAATTCGTTTTAGCCAAATATCACA 2580  
QY 2581 GCGACTGTGGGGCTCTGTTTTTTCATATATATGCGCAACCATTTCTGCGAGAGGGCT 2640  
Db 2581 GCGACTGTGGGGCTCTGTTTTTTCATATATATGCGCAACCATTTCTGCGAGAGGGCT 2640  
QY 2641 GAGTTAAATAGTGAATTAATATCTCTAACCGCGCTAATTTTACCTTAAATTTCCAT 2700  
Db 2641 GAGTTAAATAGTGAATTAATATCTCTAACCGCGCTAATTTTACCTTAAATTTCCAT 2700

Db 2641 GAGTTAAAAAGTGAATAATATCTCTAACGGCGCTAATTTTACCTTAAATTTCCCAT 2700  
 Qy 2701 GTTCCGCGGATGAGCTTTTAAATAATCAACAAAGACTTTAAACATATAAATGCAACCAATTC 2760  
 Db 2701 GTTCCGCGGATGAGCTTTTAAATAATCAACAAAGACTTTAAACATATAAATGCAACCAATTC 2760  
 Qy 2761 AATTTTCAGCTCAGACAGCAAGAGATGATTTTATGACGGGTAGCGCAGCAATGCCATC 2820  
 Db 2761 AATTTTCAGCTCAGACAGCAAGAGATGATTTTATGACGGGTAGCGCAGCAATGCCATC 2820  
 Qy 2821 AATTCACCTACACATATCCATCTCTGGCGGTAAATGTCACCTTGGTGGACAAACTCA 2880  
 Db 2821 AATTCACCTACACATATCCATCTCTGGCGGTAAATGTCACCTTGGTGGACAAACTCA 2880  
 Qy 2881 AGCAGCAGCAATACGGGGAAATATCTATCGGAGAAAGCAAAATCTTACGCTAGAAGCC 2940  
 Db 2881 AGCAGCAGCAATACGGGGAAATATCTATCGGAGAAAGCAAAATCTTACGCTAGAAGCC 2940  
 Qy 2941 AATTAACGCCCTAATCAGCAAAACATAAGGATAGAGTTATTAAGGCAATCTCACTAT 3000  
 Db 2941 AATTAACGCCCTAATCAGCAAAACATAAGGATAGAGTTATTAAGGCAATCTCACTAT 3000  
 Qy 3001 GTTAAATGGAGTTTAAATGTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTAT 3060  
 Db 3001 GTTAAATGGAGTTTAAATGTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTAT 3060  
 Qy 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACACCGGCAATTTT 3120  
 Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACACCGGCAATTTT 3120  
 Qy 3121 ACCAATAATGGCATCGCGGAAATTAATATAACACAAAGAGTGTGTAATACTTGGCAATGTT 3180  
 Db 3121 ACCAATAATGGCATCGCGGAAATTAATATAACACAAAGAGTGTGTAATACTTGGCAATGTT 3180  
 Qy 3181 ACCAATGATGTTTAAACATTAACCTACGCTAAGGCAAAAGCAAGCAATCACTC 3240  
 Db 3181 ACCAATGATGTTTAAACATTAACCTACGCTAAGGCAAAAGCAAGCAATCACTC 3240  
 Qy 3241 GCGGGAGATATATCAACAAAGAGGCTTAAATATACACAGAGTATATATGATGCT 3300  
 Db 3241 GCGGGAGATATATCAACAAAGAGGCTTAAATATACACAGAGTATATATGATGCT 3300  
 Qy 3301 GAAATCCAAATGCGGCAATATCTCGCAAAAGAGGCAACCTCAGATTTCTCCGAT 3360  
 Db 3301 GAAATCCAAATGCGGCAATATCTCGCAAAAGAGGCAACCTCAGATTTCTCCGAT 3360  
 Qy 3361 AAAATTAATATCAACAAACAGATTAACATCAAAAGGGTATGATGGAGAGGACTCTAGT 3420  
 Db 3361 AAAATTAATATCAACAAACAGATTAACATCAAAAGGGTATGATGGAGAGGACTCTAGT 3420  
 Qy 3421 TCAGATGCGCAAGTAATGCCAACCTAACTATTAAACCAAGAAATTTGAAATGACAGAA 3480  
 Db 3421 TCAGATGCGCAAGTAATGCCAACCTAACTATTAAACCAAGAAATTTGAAATGACAGAA 3480  
 Qy 3481 GACCTTAAGTATTTAGGTTTCAATTAAGAGAGATTTACAGCAAGATGGTGGAGATTTA 3540  
 Db 3481 GACCTTAAGTATTTAGGTTTCAATTAAGAGAGATTTACAGCAAGATGGTGGAGATTTA 3540  
 Qy 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCGGAGGCAAAACAGTAACCTTTTAA 3600  
 Db 3541 ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCGGAGGCAAAACAGTAACCTTTTAA 3600  
 Qy 3601 AATGTTAAAGATTAACAAATCTCTGCTACGCTCAATGTTGACACTAAATAGCAAGTG 3660  
 Db 3601 AATGTTAAAGATTAACAAATCTCTGCTACGCTCAATGTTGACACTAAATAGCAAGTG 3660  
 Qy 3661 AAAACATCTAGCAGCAATGGGGAGCTGAAAGCAATACGCAAGATACCGGCTTAAT 3720  
 Db 3661 AAAACATCTAGCAGCAATGGGGAGCTGAAAGCAATACGCAAGATACCGGCTTAAT 3720  
 Qy 3721 ATTACTGCAAAAAATGTAGAAGTAACAAAGATATTTCTCTCTCAAAACAGTAATATC 3780  
 Db 3721 ATTACTGCAAAAAATGTAGAAGTAACAAAGATATTTCTCTCTCAAAACAGTAATATC 3780

Qy 3781 ACCGCGTCGAAAGGTTACCACACAGCAGGCTCGACCATTAACGCAACAAATGCAAA 3840  
 Db 3781 ACCGCGTCGAAAGGTTACCACACAGCAGGCTCGACCATTAACGCAACAAATGCAAA 3840  
 Qy 3841 GCAAGTATTACAACCAACACAGGTGATATCAGCGGTACAGATTTCCGCTACACGCTAAGT 3900  
 Db 3841 GCAAGTATTACAACCAACACAGGTGATATCAGCGGTACAGATTTCCGCTACACGCTAAGT 3900  
 Qy 3901 GTTAGCGCGACTGGTGTATTTAAACACATAATCCGCGCTCAAAATTTGAAGCAAAATCGGT 3960  
 Db 3901 GTTAGCGCGACTGGTGTATTTAAACACATAATCCGCGCTCAAAATTTGAAGCAAAATCGGT 3960  
 Qy 3961 GAGCTTAATGTAAACAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGCTAATACGGTA 4020  
 Db 3961 GAGCTTAATGTAAACAGTGAACAGGTACAAATTTGGCGGTACAAATTTCCGCTAATACGGTA 4020  
 Qy 4021 AATGTTACGCGCAACGCTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATGCGACA 4080  
 Db 4021 AATGTTACGCGCAACGCTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATGCGACA 4080  
 Qy 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140  
 Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGGTTCTAGC 4140  
 Qy 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATCGCAGGAAGC 4200  
 Db 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATCGCAGGAAGC 4200  
 Qy 4201 ATTAATGCTGCTAATGTGACATTAATACTACAGCACTTAAACCCAGCTGGCAGGCTCG 4260  
 Db 4201 ATTAATGCTGCTAATGTGACATTAATACTACAGCACTTAAACCCAGCTGGCAGGCTCG 4260  
 Qy 4261 GATATTAAGCAACACAGCGGCACTTGGTTATTACGCAAAAGATGCTAAGCTAATCGT 4320  
 Db 4261 GATATTAAGCAACACAGCGGCACTTGGTTATTACGCAAAAGATGCTAAGCTAATCGT 4320  
 Qy 4321 GATGATCATAGGTGATGATGACAGAGTGAATGACAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380  
 Db 4321 GATGATCATAGGTGATGATGACAGAGTGAATGACAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380  
 Qy 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440  
 Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440  
 Qy 4441 AATATCATTTTCAAGAGATGTTAGAAACACTGTGCGCTTAAGAGCAAGGAAATTTGAGGTG 4500  
 Db 4441 AATATCATTTTCAAGAGATGTTAGAAACACTGTGCGCTTAAGAGCAAGGAAATTTGAGGTG 4500  
 Qy 4501 AATATATATCCAGCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCAAGGCGCTCTT 4560  
 Db 4501 AATATATATCCAGCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCAAGGCGCTCTT 4560  
 Qy 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAAACTTTGGTGTAACT 4620  
 Db 4561 GAAAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAAACTTTGGTGTAACT 4620  
 Qy 4621 GCTGTACGTTTCTGAGCCAAATTAATACAAATTTACAGTCAATACAAATTAATTTTACA 4680  
 Db 4621 GCTGTACGTTTCTGAGCCAAATTAATACAAATTTACAGTCAATACAAATTAATTTTACA 4680  
 Qy 4681 ACCAGACCGGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTCAATTTGACAAG 4740  
 Db 4681 ACCAGACCGGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTCAATTTGACAAG 4740  
 Qy 4741 GGCACGAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTCAATTTGACAAG 4800  
 Db 4741 GGCACGAGTATGTACCAATTTGCTGACGATGGACGCGGTAGTCAGTCAATTTGACAAG 4800  
 Qy 4801 GTAGATTTCACTCTGCAATGAAGTCAATTTTATTTCTGATTTTCTGTTGTTGTTTAA 4860  
 Db 4801 GTAGATTTCACTCTGCAATGAAGTCAATTTTATTTCTGATTTTCTGTTGTTGTTTAA 4860

QY 4861 GTTCAGTACGGGCTTTACCCATCTGTAAAAAATACGAGAAATACAAATAAAGTATTTT 4920  
 Db 4861 GTTCAGTACGGGCTTTACCCATCTGTAAAAAATACGAGAAATACAAATAAAGTATTTT 4920  
 QY 4921 AACAGGTTATTATTATG 4937  
 Db 4921 AACAGGTTATTATTATG 4937

RESULT 3  
 AAAS2197  
 ID AAAS2197 standard; DNA; 4940 BP.  
 AC AAAS2197;  
 XX  
 DT 11-SEP-2000 (first entry)  
 DE  
 DE Haemophilus influenzae strain 12 hmw2A gene, SEQ ID NO:70.  
 XX  
 KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;  
 KW recombinant production; Escherichia coli; antibacterial; vaccine;  
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
 KW detection; diagnosis; ds.  
 XX  
 OS Haemophilus influenzae strain 12.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 352..4785  
 FT /\*tag= a  
 FT /product= "Haemophilus influenzae strain 12 HMW2A  
 FT protein"  
 XX  
 PN WO200020609-A2.  
 XX  
 PD 13-APR-2000.  
 XX  
 XX 07-OCT-1999; 99WO-CA00938.  
 PF  
 XX 07-OCT-1998; 98US-0167568.  
 PR  
 XX 08-DEC-1998; 98US-0206942.  
 PR  
 XX (CONN-) CONNAUGHT LAB LTD.  
 PA  
 XX Loosmore SM, Yang Y, Klein MH;  
 PI WPI; 2000-303789/26.  
 XX P-PSDB; AAB01848.  
 DR  
 XX  
 XX Nucleic acid molecule for producing recombinant high molecular weight  
 PT proteins of Haemophilus which are used as a vaccine to provide  
 PT protection against Haemophilus induced diseases in humans -  
 XX  
 XX Example 16; Fig 29A-N; 307pp; English.  
 PS  
 XX The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene  
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,  
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins  
 CC and the hmwB and hmwC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMW  
 CC proteins. The modified hmwABC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMW. The invention also discloses hmwA genes (AAAS2175-A52198)  
 CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae  
 CC strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in

CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an hmwA gene from a non-typeable strain of H.  
 CC influenzae.  
 XX

SQ Sequence 4940 BP; 1729 A; 948 C; 1011 G; 1252 T; 0 Other;

Query Match 99.7%; Score 4924; DB 21; Length 4940;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 4937; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 60  
 Db 1 TAAATATACAAGATAATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA 60  
 QY 61 CACCTTTTTCAGTCTATATGCAAAATTTTAAAAAATAGTATAATCCGCATATAA 120  
 Db 61 CACCTTTTTCAGTCTATATGCAAAATTTTAAAAAATAGTATAATCCGCATATAA 120  
 QY 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTAATCTTTCATCTTTCATCTTTCAT 180  
 Db 121 AATGGTATAATCTTTTCATCTTTTCATCTTTTAATCTTTCATCTTTCATCTTTCAT 180  
 QY 181 CTTCATCTTTCATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTCATCTTTT 240  
 Db 181 CTTCATCTTTCATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTCATCTTTT 240  
 QY 241 CACATGAATGATGAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
 Db 241 CACATGAATGATGAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
 QY 301 GAAGCGAAATGATAAAGTAATTAATTTGTTCAACTTACCTTAGGAGAAATATGAACAAG 360  
 Db 301 GAAGCGAAATGATAAAGTAATTAATTTGTTCAACTTACCTTAGGAGAAATATGAACAAG 360  
 QY 361 ATATATCGTCTCAAAATTCAGCAACCGCTTGAATGCTTTGGTCTGCTGTCTGAATTTGGCA 420  
 Db 361 ATATATCGTCTCAAAATTCAGCAACCGCTTGAATGCTTTGGTCTGCTGTCTGAATTTGGCA 420  
 QY 421 CGGGTTGTGACCATTTCCACAGAAAAAGGCTTCCGCTATGTTACTATCTTTAGGTGTAC 480  
 Db 421 CGGGTTGTGACCATTTCCACAGAAAAAGGCTTCCGCTATGTTACTATCTTTAGGTGTAC 480  
 QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTACATCTATTCCA 540  
 Db 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTACATCTATTCCA 540  
 QY 541 CAATCTGTTTACCAAGCGCTTACAAGGAATGATGTAGTACAGGACGACGACGACCTATG 600  
 Db 541 CAATCTGTTTACCAAGCGCTTACAAGGAATGATGTAGTACAGGACGACGACGACCTATG 600  
 QY 601 CAAGTAGATGTTAAATAAACCATTTACCGCAACAGTGTGAGCTATCATTTAATTTGAAA 660  
 Db 601 CAAGTAGATGTTAAATAAACCATTTACCGCAACAGTGTGAGCTATCATTTAATTTGAAA 660  
 QY 661 CAATTTAAACATCGACCAAAATGAATGCTGAGTTTTTACAGAAACCAACAACCTCCGCC 720  
 Db 661 CAATTTAAACATCGACCAAAATGAATGCTGAGTTTTTACAGAAACCAACAACCTCCGCC 720  
 QY 721 GTATTCAACCGTGTATCATCTTAACCAATCTCCCAATTTAAAGGATTTTAGATTTCTAAC 780  
 Db 721 GTATTCAACCGTGTATCATCTTAACCAATCTCCCAATTTAAAGGATTTTAGATTTCTAAC 780  
 QY 781 GGACAAGTCTTTTAAACAACCCCAATTTGGTATCATCAATAGTAAAGACGCAATTTATTAA 840  
 Db 781 GGACAAGTCTTTTAAACAACCCCAATTTGGTATCATCAATAGTAAAGACGCAATTTATTAA 840  
 QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTTAACGAAACATCAAGGCGCGTAAT 900  
 Db 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTTAACGAAACATCAAGGCGCGTAAT 900







PD 29-SEP-1994.  
 XX 15-MAR-1994; 94WO-US02550.  
 XX 16-MAR-1993; 93US-0038682.  
 XX (BARE/J) BARENKAMP S J.  
 PA (SGEM/J) ST GEME J W.  
 XX Barenkamp SJ, St GEME JW;  
 PI WPI; 1994-316665/39.  
 XX P-PSDB; AAR63506.  
 DR New immunogenic high mol. wt. proteins of non typeable  
 XX Haemophilus - useful in protective vaccines  
 PT Claim 3; Page 36; 127pp; English.  
 XX  
 CC The HMW2 protein encoded by this sequence is useful in a vaccine to  
 CC protect against disease caused by non-typeable Haemophilus which are  
 CC not controlled by H. influenzae type b (Hib) vaccines, the encoded  
 CC protein can also be used as a carrier for protective Hib  
 CC polysaccharide (in a conjugate vaccine against meningitis) or for  
 CC other antigens, haptens, etc. This DNA sequence is part of a larger  
 CC sequence encoding a gene cluster for the hmw2 gene (9323 bp).  
 XX  
 SQ Sequence 4937 BP; 1728 A; 949 C; 1008 G; 1252 T; 0 other;  
 Query Match 99.5%; Score 4913; DB 15; Length 4937;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 4922; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
 QY 1 TAAATATACAGATATATAAATAAATCAAGATTTTGTGATGACAAACAAATATACAA 60  
 DB 1 TAAATATACAGATATATAAATAAATCAAGATTTTGTGATGACAAACAAATATACAA 60  
 QY 61 CACCTTTTTCGAGCTCTATGCAATATTTTAAATAAATAGTATTAATCCGCGCATATAA 120  
 DB 61 CACCTTTTTCGAGCTCTATGCAATATTTTAAATAAATAGTATTAATCCGCGCATATAA 120  
 QY 121 AATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180  
 DB 121 AATGGTATATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180  
 QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
 DB 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
 QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
 DB 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
 QY 301 GAACGCAATGATGAAGTAAATTTTAAATTTTCACTAACCTTATGAGAGGAGGAGGAGGAG 360  
 DB 301 GAACGCAATGATGAAGTAAATTTTAAATTTTCACTAACCTTATGAGAGGAGGAGGAGGAG 360  
 QY 361 ATATATGCTCTCAATTCAGCAACCGCTGAACTTTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 DB 361 ATATATGCTCTCAATTCAGCAACCGCTGAACTTTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 QY 421 CGGGGTTGTCACCATTCACAGAAAGAGGCTTCGGCTATGTTACTATCTTTAGGTGTAAC 480  
 DB 421 CGGGGTTGTCACCATTCACAGAAAGAGGCTTCGGCTATGTTACTATCTTTAGGTGTAAC 480  
 QY 481 CACTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540  
 DB 481 CACTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540  
 QY 541 CAATCTGTTTAAAGCCGCTTACAGGAATGATGATGATGATGATGATGATGATGATGATG 600  
 DB 541 CAATCTGTTTAAAGCCGCTTACAGGAATGATGATGATGATGATGATGATGATGATGATG 600

QY 601 CAACTAGATGTAATAAACCATTATCCGCAACAGCTGTTGACGCTATCATTAATTGGAAA 660  
 DB 601 CAACTAGATGTAATAAACCATTATCCGCAACAGCTGTTGACGCTATCATTAATTGGAAA 660  
 QY 661 CAATTTACATCGACCAAAATGAGTGGCAGTGTTCAGCAAGAAACAACAACTCCGCC 720  
 DB 661 CAATTTACATCGACCAAAATGAGTGGCAGTGTTCAGCAAGAAACAACAACTCCGCC 720  
 QY 721 GTATTTCAACCGTGTTCATCTCAACCAATCTCCCAATTTAAAGGATTTTAGATTCTAAC 780  
 DB 721 GTATTTCAACCGTGTTCATCTCAACCAATCTCCCAATTTAAAGGATTTTAGATTCTAAC 780  
 QY 781 GGACAAGTCTTTTAAATCAACCCAAATAGTATCAATAGTAAAGAGCGCAATTTATTAA 840  
 DB 781 GGACAAGTCTTTTAAATCAACCCAAATAGTATCAATAGTAAAGAGCGCAATTTATTAA 840  
 QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAAGGAAACATCAAGGCGCGTAAT 900  
 DB 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAAGGAAACATCAAGGCGCGTAAT 900  
 QY 901 TTCACCTTTCGAGCAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCACGGTTTAATT 960  
 DB 901 TTCACCTTTCGAGCAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCACGGTTTAATT 960  
 QY 961 ACTGTGCGTAAAGAGCGGAGTGAATCTTATGTTGCGCAAGTGAAGAGCGAGGTGTG 1020  
 DB 961 ACTGTGCGTAAAGAGCGGAGTGAATCTTATGTTGCGCAAGTGAAGAGCGAGGTGTG 1020  
 QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTCGCAGGGGCAAAAATCACCATCAGCGAT 1080  
 DB 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTCGCAGGGGCAAAAATCACCATCAGCGAT 1080  
 QY 1081 ATAATAAACCCCAACCATTTACTTTACAGCATTCGCCGCGCTGAAATTTGAAGCGGTCAATCTG 1140  
 DB 1081 ATAATAAACCCCAACCATTTACTTTACAGCATTCGCCGCGCTGAAATTTGAAGCGGTCAATCTG 1140  
 QY 1141 GCGCATATTTTTCGCAAGCGGTGAACATTAATGTCGCTGCTGCCACTATTCGAAACCAA 1200  
 DB 1141 GCGCATATTTTTCGCAAGCGGTGAACATTAATGTCGCTGCTGCCACTATTCGAAACCAA 1200  
 QY 1201 GGTAACTTTCGCTGATTTCTGTAAGCAAGATTAAGCGGCAATATTTCTTCTCCGCC 1260  
 DB 1201 GGTAACTTTCGCTGATTTCTGTAAGCAAGATTAAGCGGCAATATTTCTTCTCCGCC 1260  
 QY 1261 AAGAGGTTGAAGCGGAAATTCGCGGTGAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320  
 DB 1261 AAGAGGTTGAAGCGGAAATTCGCGGTGAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320  
 QY 1321 GGCAGCTGATGATTTACAGCGGATTAAGTCACATTAAGCAAGTTCAGTTCGACCTT 1380  
 DB 1321 GGCAGCTGATGATTTACAGCGGATTAAGTCACATTAAGCAAGTTCAGTTCGACCTT 1380  
 QY 1381 TCAGGTTAAAGAGGGGAGAACTTACCTTGGCGGTGACAGCGGCGGCAAGGTAAAC 1440  
 DB 1381 TCAGGTTAAAGAGGGGAGAACTTACCTTGGCGGTGACAGCGGCGGCAAGGTAAAC 1440  
 QY 1441 GGCATTCAATTTAGCAAGGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGC 1500  
 DB 1441 GGCATTCAATTTAGCAAGGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGC 1500  
 QY 1501 AAGAAAAAGGCGGAGCGCTATTGTTGGGCGATATTTGCGTTAATTGACGCAATATT 1560  
 DB 1501 AAGAAAAAGGCGGAGCGCTATTGTTGGGCGATATTTGCGTTAATTGACGCAATATT 1560  
 QY 1561 AAGCCTCAAGGTAGTGTGATTCGCTTAAACCGGTGTTTCTGCGAGACATCGGGGCAT 1620  
 DB 1561 AAGCCTCAAGGTAGTGTGATTCGCTTAAACCGGTGTTTCTGCGAGACATCGGGGCAT 1620  
 QY 1621 TATTTATCCANTGACAGCAATGCAATTTGTTAAACAAAGAGTGGTGTAGACCTGAT 1680  
 DB 1621 TATTTATCCANTGACAGCAATGCAATTTGTTAAACAAAGAGTGGTGTAGACCTGAT 1680  
 QY 1681 GATGTAAATTTGAAGCGGAGACCCCTTTTCGCAATTAATACCGGTATAAATGATGAATTC 1740

D <sub>b</sub>	1581	GATGTACAATTGAAGCCGAAACGCCCTTCGCAATAATACCGGTATAAATGATGAATTC	1740
Q <sub>y</sub>	1741	CCAACAGGCACCGGTGAAGCAAGCACCCTAAAAAATACGGAACCTCAAACAACGCCTA	1800
D <sub>b</sub>	1741	CCAACAGGCACCGGTGAAGCAAGCACCCTAAAAAATACGGAACCTCAAACAACGCCTA	1800
Q <sub>y</sub>	1801	ACCAATACAACTAATTTCAAATTTATCTGAAAAACGCTTGAACAATTAACGGCATCA	1860
D <sub>b</sub>	1801	ACCAATACAACTAATTTCAAATTTATCTGAAAAACGCTTGAACAATTAACGGCATCA	1860
Q <sub>y</sub>	1861	AGAAAACCTTACCGTTPAATAGCTCAATCAACATCGGAACCAACTCCCACCTTAATTCCTCAT	1920
D <sub>b</sub>	1861	AGAAAACCTTACCGTTPAATAGCTCAATCAACATCGGAACCAACTCCCACCTTAATTCCTCAT	1920
Q <sub>y</sub>	1921	AGTAAAGGTCAGCGTGCGGAGCGCTTCAGATTGATGGAGATATTACTTCTTAAGCGGGA	1980
D <sub>b</sub>	1921	AGTAAAGGTCAGCGTGCGGAGCGCTTCAGATTGATGGAGATATTACTTCTTAAGCGGGA	1980
Q <sub>y</sub>	1981	AATTTAACCATTTATCTTGGGGATGGTTCAGATTGATGTTTCATAAAAAATATTACGCTTGATCAG	2040
D <sub>b</sub>	1981	AATTTAACCATTTATCTTGGGGATGGTTCAGATTGATGTTTCATAAAAAATATTACGCTTGATCAG	2040
Q <sub>y</sub>	2041	GGTTTTTTAAATATTACCGCGCTTCGCTAGCTTTTGAAGTGGAAATACAAGCACGC	2100
D <sub>b</sub>	2041	GGTTTTTTAAATATTACCGCGCTTCGCTAGCTTTTGAAGTGGAAATACAAGCACGC	2100
Q <sub>y</sub>	2101	GACGGGGCAAAATGCTTAAATTTGTCGCCAGGGCACTGTAACCATTAACAGAGAGGCAAAA	2160
D <sub>b</sub>	2101	GACGGGGCAAAATGCTTAAATTTGTCGCCAGGGCACTGTAACCATTAACAGAGAGGCAAAA	2160
Q <sub>y</sub>	2161	GATTTTCAGGGCTAAACAACGATCTTTTAAACGGAAACGGGTAAAGTCTGTAATTCATTTCA	2220
D <sub>b</sub>	2161	GATTTTCAGGGCTAAACAACGATCTTTTAAACGGAAACGGGTAAAGTCTGTAATTCATTTCA	2220
Q <sub>y</sub>	2221	TCAGTGAATAATTTAACCCCAAACTTTAGTGGCACAATTAACATATCTGGGAATATAACA	2280
D <sub>b</sub>	2221	TCAGTGAATAATTTAACCCCAAACTTTAGTGGCACAATTAACATATCTGGGAATATAACA	2280
Q <sub>y</sub>	2281	ATTAAACCAACTACGAGAAAGAACACCTCTGTAATGGCAAAACCGCATGATTCGCACCTGG	2340
D <sub>b</sub>	2281	ATTAAACCAACTACGAGAAAGAACACCTCTGTAATGGCAAAACCGCATGATTCGCACCTGG	2340
Q <sub>y</sub>	2341	AACGTCAGTGCTCTTAATCTAGACAGCGCAAAATTTACCTTTTAAATACATTTCA	2400
D <sub>b</sub>	2341	AACGTCAGTGCTCTTAATCTAGACAGCGCAAAATTTACCTTTTAAATACATTTCA	2400
Q <sub>y</sub>	2401	AGCAATAGCAAAAGGCTTAAACAACACATAGAACTCTGAGGGGTGAATTTTAAACGGC	2460
D <sub>b</sub>	2401	AGCAATAGCAAAAGGCTTAAACAACACATAGAACTCTGAGGGGTGAATTTTAAACGGC	2460
Q <sub>y</sub>	2461	GTAATAGGCANACTGTCATTCAATCTCAAGAGAGGCGAAAGTTAATTTCAAAATFAAAA	2520
D <sub>b</sub>	2461	GTAATAGGCANACTGTCATTCAATCTCAAGAGAGGCGAAAGTTAATTTCAAAATFAAAA	2520
Q <sub>y</sub>	2521	CCAAACGAGAACATGAACAAGCAAACTTTTACCRAATTCGGTTTTTAGCCCAATATCACA	2580
D <sub>b</sub>	2521	CCAAACGAGAACATGAACAAGCAAACTTTTACCRAATTCGGTTTTTAGCCCAATATCACA	2580
Q <sub>y</sub>	2581	GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCACTTCGGCAGAGGGCT	2640
D <sub>b</sub>	2581	GCCACTGGTGGGGCTCTGTTTTTTTGTATATATATGCCAACCACTTCGGCAGAGGGCT	2640
Q <sub>y</sub>	2641	GAGTTAAAAATGAGTGAATTAATATCTCTACGGGGCTTAATTTTACCTTAAATTCOCAT	2700
D <sub>b</sub>	2641	GAGTTAAAAATGAGTGAATTAATATCTCTCTAACGGGGCTTAATTTTACCTTAAATTCOCAT	2700
Q <sub>y</sub>	2701	GTTCCGGCGGATGACGCTTTTAAAAATCAACAAGACTTAACCATTAATTCGAACCAATTC	2760
D <sub>b</sub>	2701	GTTCCGGCGGATGACGCTTTTAAAAATCAACAAGACTTAACCATTAATTCGAACCAATTC	2760
Q <sub>y</sub>	2761	AATTTACSCCTCAGACAGACGAAAGATGTTTTTATGACGGGTACGCACGCAATGCCATC	2820
D <sub>b</sub>	2761	AATTTACSCCTCAGACAGACGAAAGATGTTTTTATGACGGGTACGCACGCAATGCCATC	2820

Db	2761	AAITTCAGCCTCAGACAGACGAAGAATGATTTTATTATGACGGGTACGCACCAATGCCATC	2820
Qy	2821	AAITCAACCTACACATATCCATTTCTGGCGGTAAATGTCAACCTTGGTGACAAAACTCA	2880
Db	2821	AAITCAACCTACACATATCCATTTCTGGCGGTAAATGTCAACCTTGGTGACAAAACTCA	2880
Qy	2881	AGCAGAGCAATTACGGGGAAATATTACTATCGAAGAACGACAAATGTTACGGCTAGAGCC	2940
Db	2881	AGCAGAGCAATTACGGGGAAATATTACTATCGAAGAACGACAAATGTTACGGCTAGAGCC	2940
Qy	2941	AATAACGCCCTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGCCTC	3000
Db	2941	AATAACGCCCTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTGGCAGCTTGCCTC	3000
Qy	3001	GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGTCAGATATTAAAGGCAATCTCACTATT	3060
Db	3001	GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGTCAGATATTAAAGGCAATCTCACTATT	3060
Qy	3061	TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCCTAAATATCATCCGCGCAATTTT	3120
Db	3061	TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCCTAAATATCATCCGCGCAATTTT	3120
Qy	3121	ACCAATAATGGCACTGCCGAAATTAATATAACACAAGGAGTGTAAAACTTGGCAATGTT	3180
Db	3121	ACCAATAATGGCACTGCCGAAATTAATATAACACAAGGAGTGTAAAACTTGGCAATGTT	3180
Qy	3181	ACCAATGATGGTATTAACATATACCACCTCAGCTAAACGCAACCAAGACATCATC	3240
Db	3181	ACCAATGATGGTATTAACATATACCACCTCAGCTAAACGCAACCAAGACATCATC	3240
Qy	3241	GGCGGAGATTAATCAACAAAAAGGAAGCTTAAATATTACAGACAGTAATAATGATGCT	3300
Db	3241	GGCGGAGATTAATCAACAAAAAGGAAGCTTAAATATTACAGACAGTAATAATGATGCT	3300
Qy	3301	GAATCCCAATTTGGCGCAATATCTCGCAAAAAGRAGGCAACCTCACGATTTCTCCGAT	3360
Db	3301	GAATCCCAATTTGGCGCAATATCTCGCAAAAAGRAGGCAACCTCACGATTTCTCCGAT	3360
Qy	3361	AAAAATTAATATCACCANAACAGATAACAATCAAAAAGGTTATGTGGAGAGGACTCTAGT	3420
Db	3361	AAAAATTAATATCACCANAACAGATAACAATCAAAAAGGTTATGTGGAGAGGACTCTAGT	3420
Qy	3421	TCAGATGGCAAGTAATGCCAACCTAACTATTAAACCAAGAATTTGAAATTCAGAAA	3480
Db	3421	TCAGATGGCAAGTAATGCCAACCTAACTATTAAACCAAGAATTTGAAATTCAGAAA	3480
Qy	3481	GACCTAAATATTTCAGGTTTCAATTAAGACAGATTACAGCCAAAAGATGGTAGAGATTTA	3540
Db	3481	GACCTAAATATTTCAGGTTTCAATTAAGACAGATTACAGCCAAAAGATGGTAGAGATTTA	3540
Qy	3541	ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAGGCCAAACAGTAACCTTTAAC	3600
Db	3541	ACTATTGGCAACAGTAATGACGGTAACAGCGGTGCCGAGGCCAAACAGTAACCTTTAAC	3600
Qy	3601	AATGTTAAAGATTCAAAAATCTCTGCTCAGCGTCAATGTGACACTAAATAGCAAACTG	3660
Db	3601	AATGTTAAAGATTCAAAAATCTCTGCTCAGCGTCAATGTGACACTAAATAGCAAACTG	3660
Qy	3661	AAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCGAACAGTACCGGCTTAACT	3720
Db	3661	AAACATCTAGCAGCAATGGCGGACGTGAAAGCAATAGCGAACAGTACCGGCTTAACT	3720
Qy	3721	ATTACTGCAAAAATGTAGAAGTAACCAAGATATTACTTCTCTCAAAACAGTAATATC	3780
Db	3721	ATTACTGCAAAAATGTAGAAGTAACCAAGATATTACTTCTCTCAAAACAGTAATATC	3780
Qy	3781	ACCOCGTGCGAAAAAGGTTACCCACACAGCGCTCAGCAATTAACGCAACAAATGGCAAA	3840
Db	3781	ACCOCGTGCGAAAAAGGTTACCCACACAGCGCTCAGCAATTAACGCAACAAATGGCAAA	3840
Qy	3841	GCAAGTATTCAACCAACAAACNGGTGATATACGGCGTACGATTTCCCGGTAAACACGGTAAGT	3900
Db	3841	GCAAGTATTCAACCAACAAACNGGTGATATACGGCGTACGATTTCCCGGTAAACACGGTAAGT	3900

```
QY 3901 GTTAGCGGACTGGTGAATTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960
DB 3901 GTTAGCGGACTGGTGAATTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960
QY 3961 GAGGCTAATGTAAACAGTCAACAGAGTCAAAATTTGGCGGTACAAATTTCCGGTAAATACGGTA 4020
DB 3961 GAGGCTAATGTAAACAGTCAACAGAGTCAAAATTTGGCGGTACAAATTTCCGGTAAATACGGTA 4020
QY 4021 AATGTTACGGCAAAACGCTGGCGATTTAAACAGTTGGGAATGGCGCAGAAATTAATCGGACA 4080
DB 4021 AATGTTACGGCAAAACGCTGGCGATTTAAACAGTTGGGAATGGCGCAGAAATTAATCGGACA 4080
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGCATCTACTGAAGCCGGTTCTAGC 4140
DB 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGCATCTACTGAAGCCGGTTCTAGC 4140
QY 4141 ATCACTTCAACTAAGGCTAGGTAGACCTCTGGCTCAGAAATGGTAGCATCGCAGGAAGC 4200
DB 4141 ATCACTTCAACTAAGGCTAGGTAGACCTCTGGCTCAGAAATGGTAGCATCGCAGGAAGC 4200
QY 4201 ATTAATGCTGCTAATGTGACATTAATACTACAGGCACTTTAACCACTGGCGAGGCTCG 4260
DB 4201 ATTAATGCTGCTAATGTGACATTAATACTACAGGCACTTTAACCACTGGCGAGGCTCG 4260
QY 4261 GATATTAAAGCAACAGCGGCACCTTTGGTTTATTAAAGCAAAAGATGCTAAGCTTAAATGGT 4320
DB 4261 GATATTAAAGCAACAGCGGCACCTTTGGTTTATTAAAGCAAAAGATGCTAAGCTTAAATGGT 4320
QY 4321 GATGCATCAGTGATAGTACAGAAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
DB 4321 GATGCATCAGTGATAGTACAGAAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG 4380
QY 4381 ACTGGGCAACCTCAACGAGTGTGAATATCACTGGGGATTTTAAACACAGTAAATGGGTTA 4440
DB 4381 ACTGGGCAACCTCAACGAGTGTGAATATCACTGGGGATTTTAAACACAGTAAATGGGTTA 4440
QY 4441 AATATCATTTTGAAGATGTTAGAAACACCTGTGCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500
DB 4441 AATATCATTTTGAAGATGTTAGAAACACCTGTGCGCTTAAAGAGCAAGGAAATTTGAGGTG 4500
QY 4501 AAATATATCCAGCAGGTGTAGCAAGTGTGAAGCAAGTAAATTTGAAGCGCAAGGAAATTTGAGGTG 4560
DB 4501 AAATATATCCAGCAGGTGTAGCAAGTGTGAAGCAAGTAAATTTGAAGCGCAAGGAAATTTGAGGTG 4560
QY 4561 GAAAAGTAAAGATTTATCTGTATGAGAAAGAGAAACATTTAGCTAACTTTGGTTAAGT 4620
DB 4561 GAAAAGTAAAGATTTATCTGTATGAGAAAGAGAAACATTTAGCTAACTTTGGTTAAGT 4620
QY 4621 GCTGTACGTTTGTGTAGCCAAATTAATAACAATTTACAGTCAATACACAAATGAAATTTACA 4680
DB 4621 GCTGTACGTTTGTGTAGCCAAATTAATAACAATTTACAGTCAATACACAAATGAAATTTACA 4680
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAAGTAAAGCGCTGTTCTCAAGTGGTAAT 4740
DB 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAAGTAAAGCGCTGTTCTCAAGTGGTAAT 4740
QY 4741 GCGCACGAGTATGTACCAATTTGCTGCATGAGCAAGCGGTAGTCAAGTAAATTTGACAAG 4800
DB 4741 GCGCACGAGTATGTACCAATTTGCTGCATGAGCAAGCGGTAGTCAAGTAAATTTGACAAG 4800
QY 4801 GTAGATTTTCATCTCAGTGAAGTCAATTTTATTTCTGATTTATTTACTGTCTGGGTTAAA 4860
DB 4801 GTAGATTTTCATCTCAGTGAAGTCAATTTTATTTCTGATTTATTTACTGTCTGGGTTAAA 4860
QY 4861 GTTCAGTACGGGCTTTACCCATCTTGTGTAATAAATTTACGAGAAATACAATAAGATTTT 4920
DB 4861 GTTCAGTACGGGCTTTACCCATCTTGTGTAATAAATTTACGAGAAATACAATAAGATTTT 4920
QY 4921 AACAGGTTATTATG 4937
DB 4921 AACAGGTTATTATG 4937
```

```
RESULT 5
AAQ49509 standard; DNA; 9323 BP.
XX AAQ49509;
XX 26-APR-1994 (first entry)
XX Gene cluster for high molecular weight protein 2 (HMW2).
XX HMW; high molecular weight protein; virus; vaccine; influenza;
XX epitope; immunity; haemophilus influenzae; gene cluster; ss.
XX Haemophilus influenzae.
XX Key Location/Qualifiers
XX CDS 792..5222
XX /product= "High molecular weight protein 2"
XX /tag= a
XX CDS 5375..7009
XX /tag= b
XX /note= "One of a gene cluster for High molecular"
XX 7149..9098
XX /tag= c
XX /note= "One of a gene cluster for High molecular"
XX WO9319090-A.
XX 30-SEP-1993.
XX 16-MAR-1993; 93WO-US02166.
XX 16-MAR-1992; 92GB-0005704.
XX (BARE/) BARENKAMP S J.
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX Barenkamp SJ;
XX WPI; 1993-320683/40.
XX P-PSDB; AAR41728, AAR41729, AAR41730.
XX High molecular weight surface proteins - of non-typeable
XX haemophilus which exhibit immunogenic properties
XX Claim 10; Figure 7; 100pp; English.
XX The isolation and purification of the high molecular weight protein
XX enables the identification of the major protective epitopes of the
XX protein by conventional epitope mapping. These epitopes can then be
XX synthesised using standard techniques and incorporated into fully
XX synthetic or recombinant vaccines.
XX Sequence 9323 BP; 3133 A; 1829 C; 1833 G; 2528 T; 0 other;
```

```
Query Match 97.8%; Score 4828.2; DB 14; Length 9323;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 4869; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 1 TAAATATACAGATTAATAAATAAATCAAGATTTTGTGATGACAAACAAATTTACAA 60
DB 441 TAAATATACAGATTAATAAATAAATCAAGATTTTGTGATGACAAACAAATTTACAA 500
QY 61 CACCTTTTTCAGCTCTATATGCAATATTTTAAAAAATAGTATAATCCGCATATAA 120
DB 501 CACCTTTTTCAGCTCTATATGCAATATTTTAAAAAATAGTATAATCCGCATATAA 560
QY 121 AATGGTAAATCTTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAAT 180
DB 561 AATGGTAAATCTTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAAT 620
```







QY 4561 GAAAAAGTAAAGATTTATCTGATCAAGAGAGAGAAACATAGCTAAACTTGGTGAAGT 4620  
 DB 5001 GAAAAAGTAAAGATTTATCTGATCAAGAGAGAGAAACATAGCTAAACTTGGTGAAGT 5060  
 QY 4621 GCTGTAGCTTTTGTGAGCCAAATATACAAATACAGTCAATACACAAATGAATTTACA 4680  
 DB 5061 GCTGTAGCTTTTGTGAGCCAAATATACAAATACAGTCAATACACAAATGAATTTACA 5120  
 QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGTGTCTTCTCAAGTGGTAAT 4740  
 DB 5121 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGTGTCTTCTCAAGTGGTAAT 5180  
 QY 4741 GGGCGACGAGTATGATCAATGTTGCTGACGATGGACGCGTAGTCAGTAAATGACAAG 4800  
 DB 5181 GGGCGACGAGTATGATCAATGTTGCTGACGATGGACGCGTAGTCAGTAAATGACAAG 5240  
 QY 4801 GTAGATTTTCATCTGCAATGAAGTCAATTTTATTTTTCGTAATTTTACTGTGTGGTTAAA 4860  
 DB 5241 GTAGATTTTCATCTGCAATGAAGTCAATTTTATTTTTCGTAATTTTACTGTGTGGTTAAA 5300  
 QY 4861 GTTCAGTACGGCTTTACCCATCTTGTAAATAATACGAGAAATACAAATAAATTTT 4920  
 DB 5301 GTTCAGTACGGCTTTACCCATCTTGTAAATAATACGAGAAATACAAATAAATTTT 5360  
 QY 4921 AACAGGTTATTTATG 4937  
 DB 5361 AACAGGTTATTTATG 5377

RESULT 6  
 ID AAT90997 standard; DNA; 9323 BP.  
 XX AC AAT90997;  
 XX DT 14-APR-1998 (first entry)  
 XX DE Non-typeable Haemophilus influenzae hmw2 gene cluster.  
 KW Non-typeable Haemophilus; high molecular weight surface protein;  
 KW Hmw2; hmw2 gene; immunogen; vaccine; otitis media; ss.  
 XX OS Haemophilus influenzae strain 12.  
 XX FH Key Location/Qualifiers  
 FT CDS 792..5222  
 FT /\*tag= a  
 FT /note= "hmw2 gene"  
 FT CDS 5375..7009  
 FT /\*tag= b  
 FT /note= "ORF-b"  
 FT CDS 7249..9198  
 FT /\*tag= c  
 FT /note= "ORF-c"  
 XX WO9736914-A1.  
 XX 09-OCT-1997.  
 XX 01-APR-1997; 97WO-US04707.  
 XX 01-APR-1996; 96US-0617697.  
 XX (BARE/) BARENKAMP S J.  
 XX Barenkamp SJ;  
 XX PI  
 DR WPI; 1997-503038/46.  
 DR P-PSDB; AAW30294.  
 XX High molecular weight proteins of non-typeable Haemophilus  
 PT influenzae - useful for vaccine production

XX Disclosure; Page 82-87; 183pp; English.  
 XX This nucleic acid comprises a gene cluster for the hmw2 gene (see  
 CC also AAT90995) that encodes high molecular weight surface protein  
 CC Hmw2 (see AAW30294) of non-typeable Haemophilus influenzae strain 12.  
 CC In addition to the hmw2 gene there are 2 additional downstream  
 CC open reading frames that are required for the correct processing  
 CC and secretion of the hmw1 gene product. The ORF-b derived amino  
 CC acid sequence demonstrates similarity with the derived amino acid  
 CC sequences of 2 genes which encode proteins required for secretion  
 CC and activation of haemolysins of P. mirabilis and S. marcescens.  
 CC Hmw proteins (see AAW30291-94) can be used in vaccines, as immunogens  
 CC for preparation of antibodies and as antigens for detection of  
 CC these antibodies.  
 XX Sequence 9323 BP; 3135 A; 1829 C; 1831 G; 2528 T; 0 other;  
 SQ Query Match 97.6%; Score 4820.2; DB 18; Length 9323;  
 Best Local Similarity 98.5%; Pred. No. 0;  
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
 QY 1 TAAATATACAGATATAATAAATAAATCAAGATTTTGTGATGACAAACAATTTACAA 60  
 DB 441 TAAATATACAGATATAATAAATAAATCAAGATTTTGTGATGACAAACAATTTACAA 500  
 QY 61 CACCTTTTTCAGTCTATATGCAAAATATTTTAAATAAATAGTATAAATCCGCATATAA 120  
 DB 501 CACCTTTTTCAGTCTATATGCAAAATATTTTAAATAAATAGTATAAATCCGCATATAA 560  
 QY 121 AATGGTATAATCTTTCATCTTTCAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180  
 DB 561 AATGGTATAATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620  
 QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
 DB 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680  
 QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
 DB 681 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740  
 QY 301 GAACGCAATGATAAGTAATTTAATTTGTTCAACTAACCTTAGGAGAGAAATATGAACAAG 360  
 DB 741 GAACGCAATGATAAGTAATTTAATTTGTTCAACTAACCTTAGGAGAGAAATATGAACAAG 800  
 QY 361 ATATATCGTCTCAAAATTCAGCAACCGCTTGAATGCTTTGGTTGCTGCTGAATTTGGCA 420  
 DB 801 ATATATCGTCTCAAAATTCAGCAACCGCTTGAATGCTTTGGTTGCTGCTGAATTTGGCA 860  
 QY 421 CGGGTTGTGACCAATTCACAGAAAAAGGCTTCCGCTATGTTTACTATCTTTTAGGTGTAAC 480  
 DB 861 CGGGTTGTGACCAATTCACAGAAAAAGGCTTCCGCTATGTTTACTATCTTTTAGGTGTAAC 920  
 QY 481 CACTTAGCGTTAAAGCCACTTTCGCTATGTTTACTATCTTTTAGGTGTAACCTTACATCTAT 540  
 DB 921 CACTTAGCGTTAAAGCCACTTTCGCTATGTTTACTATCTTTTAGGTGTAACCTTACATCTAT 980  
 QY 541 CAATCTGTTTGTAGCAAGCGCTTACAGAAATGAGTGTAGTACAGCGCACAGCCACTATG 600  
 DB 981 CAATCTGTTTGTAGCAAGCGCTTACAGAAATGAGTGTAGTACAGCGCACAGCCACTATG 1040  
 QY 601 CAAGTATAGTGAATAAACCATTATCCGCAACAGTGTGTAGCGCTATCATTTAATTTGAAA 660  
 DB 1041 AAGAAAAACAAGTAATAAACCATTATCCGCAACAGTGTGTAGCGCTATCATTTAATTTGAAA 1100  
 QY 661 CAATTTAACATCGACCAAAATGAATGGTGCAGTTTTACAGAAACAACAACCTCCGCC 720  
 DB 1101 CAATTTAACATCGACCAAAATGAATGGTGCAGTTTTACAGAAACAACAACCTCCGCC 1160  
 QY 721 GTATTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTTAAC 780  
 DB 1161 GTATTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGGATTTTAGATTCTTAAC 1220

Qy 781 GGACAGCTCTTTTAAATCAACCCAAAGGTTATCACAAATAGGTAAAGACGCAATTTAAAC 840  
Db 1221 GGACAGCTCTTTTAAATCAACCCAAAGGTTATCACAAATAGGTAAAGACGCAATTTAAAC 1280  
Qy 841 ACTAATGGCTTACGGCTTCTACGCTAGACATTTCTAAGCAAAACATCAAGGCGCTAAT 900  
Db 1281 ACTAATGGCTTACGGCTTCTACGCTAGACATTTCTAAGCAAAACATCAAGGCGCTAAT 1340  
Qy 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTCTGAATCACGGTCTTAAAT 960  
Db 1341 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTCTGAATCACGGTCTTAAAT 1400  
Qy 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGGTGGCAAGGTGAAACAGAGGCTGTG 1020  
Db 1401 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGGTGGCAAGGTGAAACAGAGGCTGTG 1460  
Qy 1021 ATTAGCTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATTCACCATCAGCGAT 1080  
Db 1461 ATTAGCTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATTCACCATCAGCGAT 1520  
Qy 1081 ATATTAACCCCAACCATTTACTTACAGCATTTCCCGCGCTGAAATTAAGCGGCTCAATCTG 1140  
Db 1521 ATATTAACCCCAACCATTTACTTACAGCATTTCCCGCGCTGAAATTAAGCGGCTCAATCTG 1580  
Qy 1141 GCGGATATTTTGGCCAAAGGGGTAACATTAATGTCGCTGCTGCGCAGTATTCGAACCAA 1200  
Db 1581 GCGGATATTTTGGCCAAAGGGGTAACATTAATGTCGCTGCTGCGCAGTATTCGAACCAA 1640  
Qy 1201 GGTAAACTTTCTGCTGATTTCTTAAGCAAAAGATATAAGCGCGCAATATTTGTTCTTCCGCC 1260  
Db 1641 GGTAAACTTTCTGCTGATTTCTTAAGCAAAAGATATAAGCGCGCAATATTTGTTCTTCCGCC 1700  
Qy 1261 AAAGAGGGTGAAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1320  
Db 1701 AAAGAGGGTGAAGCGGAAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1760  
Qy 1321 GGCAGCTGATGATTACAGGCGATTAAGTCAATTAACAGAGGTGCGAGTTATCGACCTT 1380  
Db 1761 GGCAGCTGATGATTAAAGTCCGATTAAGTCAATTAACAGAGGTGCGAGTTATCGACCTT 1820  
Qy 1381 TCAGGTAAGAGGGGGAACCTTACCTTCGCGGTGACGAGCGCGCGAGGTAAAC 1440  
Db 1821 TCAGGTAAGAGGGGGAACCTTACCTTCGCGGTGACGAGCGCGCGAGGTAAAC 1880  
Qy 1441 GGCATTCATTTAGCAAAAGAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500  
Db 1881 GGCATTCATTTAGCAAAAGAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1940  
Qy 1501 AAAGAAAAGGGGACGCGCTATTGTGGGGCGATATTGGCGTTAAATGACGGCAATTT 1560  
Db 1941 AAAGAAAAGGGGACGCGCTATTGTGGGGCGATATTGGCGTTAAATGACGGCAATTT 2000  
Qy 1561 AACGCTCAAGGTAGTGGTATATCGCTAAAACCGGTGTTTGTGGAGACATCGGGGCAT 1620  
Db 2001 AACGCTCAAGGTAGTGGTATATCGCTAAAACCGGTGTTTGTGGAGACATCGGGGCAT 2060  
Qy 1621 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTCGTAGACCCCTGAT 1680  
Db 2061 TATTTATCCATTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTTCGTAGACCCCTGAT 2120  
Qy 1681 GATGTAACAAATTAAGCGGAGACCCCTTCGCAATATACCGGTATTAATGATCAATTC 1740  
Db 2121 GATGTAACAAATTAAGCGGAGACCCCTTCGCAATATACCGGTATTAATGATCAATTC 2180  
Qy 1741 CCAACAGGCAAGGGTGAAGCAAGGACCCCTTAAAAAATAGGCACTCAAAACACGCTA 1800  
Db 2181 CCAACAGGCAAGGGTGAAGCAAGGACCCCTTAAAAAATAGGCACTCAAAACACGCTA 2240  
Qy 1801 ACCAATFACAACATTTTCAAAATTTATCTGAAAAACGCGCTGGACAATGAATATAACGGCATCA 1860  
Db 2241 ACCAATFACAACATTTTCAAAATTTATCTGAAAAACGCGCTGGACAATGAATATAACGGCATCA 2300

Qy 1861 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCACCTTAAATCTCCAT 1920  
Db 2301 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCAACTCCACCTTAAATCTCCAT 2360  
Qy 1921 AGTAAAGTTCAGGCTGGGAGGCGTTTCAAGTTGATGAGATATTTACTTCTTAAAGCGCA 1980  
Db 2361 AGTAAAGTTCAGGCTGGGAGGCGTTTCAAGTTGATGAGATATTTACTTCTTAAAGCGCA 2420  
Qy 1981 AATTTAAACCATTTATTCGCGGATGGGTGATGTTTCATFAAAAAATTTACGGTTGATCAG 2040  
Db 2421 AATTTAAACCATTTATTCGCGGATGGGTGATGTTTCATFAAAAAATTTACGGTTGATCAG 2480  
Qy 2041 GGTTTTTTAAATATTTACCGCGCTTCCGCTAGCTTTTGAAGGTGGAATAAACAACGACGC 2100  
Db 2481 GGTTTTTTAAATATTTACCGCGCTTCCGCTAGCTTTTGAAGGTGGAATAAACAACGACGC 2540  
Qy 2101 GACGGGCAAAATGCTAAAAATTTGTCGCCAGGGCACTGTAAACCATTTACAGGAGGAGGAAAA 2160  
Db 2541 GACGGGCAAAATGCTAAAAATTTGTCGCCAGGGCACTGTAAACCATTTACAGGAGGAGGAAAA 2600  
Qy 2161 GATTTTCAGGCTTAAACAACGCTATCTTTAAACGGAACGGGTAAAAGGCTGTAATATCATTTCA 2220  
Db 2601 GATTTTCAGGCTTAAACAACGCTATCTTTAAACGGAACGGGTAAAAGGCTGTAATATCATTTCA 2660  
Qy 2221 TCAGTCAATTAATTTAACCCACATCTTACTGTCACAAATTAACATATCTCGGAATATAACA 2280  
Db 2661 TCAGTCAATTAATTTAACCCACATCTTACTGTCACAAATTAACATATCTCGGAATATAACA 2720  
Qy 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTTGCAACCAACCATGATTCGCACTGG 2340  
Db 2721 ATTAACCAAACTACGAGAAAGAACACCTCGTATTTGCAACCAACCATGATTCGCACTGG 2780  
Qy 2341 AACGTCACTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTATTAATACATTTCA 2400  
Db 2781 AACGTCACTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTATTAATACATTTCA 2840  
Qy 2401 AGCAATAGCAAAAGGCTTAAACACACAGTATAGAGCTCTGAGGGGTGAATTTTAAACGCG 2460  
Db 2841 AGCAATAGCAAAAGGCTTAAACACACAGTATAGAGCTCTGAGGGGTGAATTTTAAACGCG 2900  
Qy 2461 GTAAATGGCAACATGTCTTCAATCTCAAAAGAGGAGCGAAAGTTTAAATTTCAATTTAAA 2520  
Db 2901 GTAAATGGCAACATGTCTTCAATCTCAAAAGAGGAGCGAAAGTTTAAATTTCAATTTAAA 2960  
Qy 2521 CCAACGAGAACATGAACACAAACACCTTTTACCAATTCGCTTTTAAAGCAATATCACA 2580  
Db 2961 CCAACGAGAACATGAACACAAACACCTTTTACCAATTCGCTTTTAAAGCAATATCACA 3020  
Qy 2581 GCCACTGGTGGGCTCTGTTTGTGATATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
Db 3021 GCCACTGGTGGGCTCTGTTTGTGATATATATGCCAACCATTTCTGGCAGAGGGCT 3080  
Qy 2641 GAGTTAAATATGAGTGAATTAATATCTCTAACGGGCTTAATTTTACCTTAAATTTCCCAT 2700  
Db 3081 GAGTTAAATATGAGTGAATTAATATCTCTAACGGGCTTAATTTTACCTTAAATTTCCCAT 3140  
Qy 2701 GTTCGCGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATTAATGCAACCAATTC 2760  
Db 3141 GTTCGCGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATTAATGCAACCAATTC 3200  
Qy 2761 AATTTACGCTCTACAGACAGAAAGATGATTTTATGACGGGTACGCAACCAATGCGCATC 2820  
Db 3201 AATTTACGCTCTACAGACAGAAAGATGATTTTATGACGGGTACGCAACCAATGCGCATC 3260  
Qy 2821 AATTTCAACCTACACATATCCATTTCTGGGCGTAAATGTCACCTTGTGGACAAAACCTCA 2880  
Db 3261 AATTTCAACCTACACATATCCATTTCTGGGCGTAAATGTCACCTTGTGGACAAAACCTCA 3320  
Qy 2881 AGCAGCAGCATTCGCGGATATTTACTATCGAGAAAGCAGCAATGTTAGCTAGAGCC 2940  
Db 3321 AGCAGCAGCATTCGCGGATATTTACTATCGAGAAAGCAGCAATGTTAGCTAGAGCC 3380  
Qy 2941 AATAACGCGCCCTAATCAGCAAAACATTAAGGGATAGAGTTATAAAACCTTGGCAGCTTGCTC 3000

```
|||||
Db 3381 AATAACGCCCTAATCAGCAAAACATAAGGATAGACTTATANAACCTGGCAGCTTGCCTC 3440
QY 3001 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTATT 3060
Db 3441 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCGAGATATTAAGGCAATCTCACTATT 3500
QY 3061 TCAGAAAGGCCACCTTTTAAAGGAAGACACTAGAGATACCCCTAAATATCACCGGCAATTTT 3120
Db 3501 TCAGAAAGGCCACCTTTTAAAGGAAGACACTAGAGATACCCCTAAATATCACCGGCAATTTT 3560
QY 3121 ACCAATTAATGGCACTGCCGAAATTAATATAACACAAAGGAGTGGTAAACCTTGGCAATGTT 3180
Db 3561 ACCAATTAATGGCACTGCCGAAATTAATATAACACAAAGGAGTGGTAAACCTTGGCAATGTT 3620
QY 3181 ACCAATGATGGTGATTTAAACATTTACCCTACCGCTTAAGGCAACCAAGGAGTGGTAAACCTTGGCAATGTT 3680
Db 3621 ACCAATGATGGTGATTTAAACATTTACCCTACCGCTTAAGGCAACCAAGGAGTGGTAAACCTTGGCAATGTT 3740
QY 3241 GCGGAGATATATAACACAAAGGAGTGGTAAATATTAACAGACAGTAAATATGATGCT 3300
Db 3681 GCGGAGATATATAACACAAAGGAGTGGTAAATATTAACAGACAGTAAATATGATGCT 3740
QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGGAGTGGTAAACCTTGGCAATGTT 3360
Db 3741 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGGAGTGGTAAACCTTGGCAATGTT 3800
QY 3361 AAAATTAATATCACCAAAACAGATACCAATCAAAAGGAGTGGTAAACCTTGGCAATGTT 3420
Db 3801 AAAATTAATATCACCAAAACAGATACCAATCAAAAGGAGTGGTAAACCTTGGCAATGTT 3480
QY 3421 TCAGATGCCACAAGTAAATGCCAACCTTAATTAACCAAGAGTGGTAAACCTTGGCAATGTT 3480
Db 3861 TCAGATGCCACAAGTAAATGCCAACCTTAATTAACCAAGAGTGGTAAACCTTGGCAATGTT 3920
QY 3481 GACCTAAAGTATTTCAAGTTTCAATTAAGCAGAGATTTACAGCCAAAGAGTGGTAAACCTTGGCAATGTT 3540
Db 3921 GACCTAAAGTATTTCAAGTTTCAATTAAGCAGAGATTTACAGCCAAAGAGTGGTAAACCTTGGCAATGTT 3980
QY 3541 ACTATTTGGCAACAGTAAATGACGGTAAACAGCGGTGCGGAGGCAAGCAAGTAACTTTTAACT 3600
Db 3981 ACTATTTGGCAACAGTAAATGACGGTAAACAGCGGTGCGGAGGCAAGCAAGTAACTTTTAACT 3660
QY 3601 AATGTTAAAGATTTCAAAATCTCTGCTGACGGTCAATGTCACACTTAATAGCAAGTAACT 4000
Db 4041 AATGTTAAAGATTTCAAAATCTCTGCTGACGGTCAATGTCACACTTAATAGCAAGTAACT 4060
QY 3661 AANAATCTAGCAGCAATGGCGGAGCTGAAAGCAATAGCAGCAAGTAACTTAACTTAACT 4100
Db 4101 AANAATCTAGCAGCAATGGCGGAGCTGAAAGCAATAGCAGCAAGTAACTTAACTTAACT 4160
QY 3721 ATTACTGCAAAATGTAAGTAACAAAGATATTAATCTCTCAAAACAGTAAATATC 3780
Db 4161 ATTACTGCAAAATGTAAGTAACAAAGATATTAATCTCTCAAAACAGTAAATATC 4220
QY 3781 ACCGGCTCGGAAAGGTTTACCACCAACAGCGCTGCGACATTAAGCAACCAATGGCAAA 3840
Db 4221 ACCGGCTCGGAAAGGTTTACCACCAACAGCGCTGCGACATTAAGCAACCAATGGCAAA 4280
QY 3841 GCAAGTATTAACCAACCAAGGTAATATACGGGTACGATTTCCGGTAAACCGTAACT 3900
Db 4281 GCAAGTATTAACCAACCAAGGTAATATACGGGTACGATTTCCGGTAAACCGTAACT 4340
QY 3901 GTTAGCGGACTGGTGATTTAACCACCTAAATCCCGCTCAAAATTAAGCAAAATCGGGT 3960
Db 4341 GTTAGCGGACTGGTGATTTAACCACCTAAATCCCGCTCAAAATTAAGCAAAATCGGGT 4000
QY 3961 GAGGCTAATGTAACAGTGCACAGGTAATATGCGGTACAAATTTCCGGTAAATACGGTA 4020
Db 4401 GAGGCTAATGTAACAGTGCACAGGTAATATGCGGTACAAATTTCCGGTAAATACGGTA 4460
QY 4021 AATGTTACGGCAACGCTGGGATTTACAGTTCGGGATTCGGGAGTAAATTAATGCGACA 4080
|||||
```

```
Db 4461 AATGTTACGGCAAAACGCTGGCGATTTAACAGTTGGGAATGCGCAGAAATTAATGCGACA 4520
QY 4081 GAAGGAGCTGCAACCTTTAAACCGCAACAGGAATACCTTTGACTACTGAAGCCGGTTCTAGC 4140
Db 4521 GAAGGAGCTGCAACCTTTAAACCGCAACAGGAATACCTTTGACTACTGAAGCCGGTTCTAGC 4580
QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCGCAGGAAGC 4200
Db 4581 ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCGCAGGAAGC 4640
QY 4201 ATTAATGCTCTTAATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Db 4641 ATTAATGCTCTTAATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4700
QY 4261 GATATTAAGCAACACCGCGACCTTTGGTTTAAACCAAAAGATGCTAAAGTAAATGGT 4320
Db 4701 GATATTAAGCAACACCGCGACCTTTGGTTTAAACCAAAAGATGCTAAAGTAAATGGT 4760
QY 4321 GATGATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACCGAAGCGCTCTGCTAGTGTG 4380
Db 4761 GATGATCAGGTGATAGTACAGAAAGTGAATGCAAGTCAACCGAAGCGCTCTGCTAGTGTG 4820
QY 4381 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGTGA 4440
Db 4821 ACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGTGA 4880
QY 4441 AATATCATTTTCCGAAAGTGGTAGAAACACTGTGCGCTTAAGAGCAAGGAATTTAGGTTG 4500
Db 4881 AATATCATTTTCCGAAAGTGGTAGAAACACTGTGCGCTTAAGAGCAAGGAATTTAGGTTG 4940
QY 4501 AATATATATCAGCGCAGGTGTAGCAAGTGTAGAAAGAGTAAATTCAGCGAAGCAACCGCTT 4560
Db 4941 AATATATATCAGCGCAGGTGTAGCAAGTGTAGAAAGAGTAAATTCAGCGAAGCAACCGCTT 5000
QY 4561 GAAAGAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACTTGGTGAAGT 4620
Db 5001 GAAAGAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACTTGGTGAAGT 5060
QY 4621 GCTGACGCTTTTGTGAGCAAAATTAATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
Db 5061 GCTGACGCTTTTGTGAGCAAAATTAATACAAATTTACAGTCAATACACAAATGAATTTACA 5120
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGCTGTTTCTCAAGTGGTAAAT 4740
Db 5121 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGCTGTTTCTCAAGTGGTAAAT 5180
QY 4741 GCGCAGCAGATGTACCAATCTTCTGACAGTGGACACCGCTAGTCACTAATTTGACAAG 4800
Db 5181 GCGCAGCAGATGTACCAATCTTCTGACAGTGGACACCGCTAGTCACTAATTTGACAAG 5240
QY 4801 GTAGATTTTCACTCTGCAATGAAGTCAATTTTATTTTCTGATTTTACTCTGTTGGGTAAA 4860
Db 5241 GTAGATTTTCACTCTGCAATGAAGTCAATTTTATTTTCTGATTTTACTCTGTTGGGTAAA 5300
QY 4861 GTTACGACGGCTTTTACCCTCTTTGTAATAAATTAACGAGAAATACAAATTAATTTT 4920
Db 5301 GTTACGACGGCTTTTACCCTCTTTGTAATAAATTAACGAGAAATACAAATTAATTTT 5360
QY 4921 AACAGGTTTATTTATG 4937
Db 5361 AACAGGTTTATTTATG 5377
```

RESULT 7

AAA52198

ID AAA52198 standard; DNA; 3263 BP.

XX AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;

XX AC AAA52198;



KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;  
 KW recombinant production; Escherichia coli; antibacterial; vaccine;  
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
 KW detection; diagnosis; ds.

XX Haemophilus influenzae strain 12.

XX Key Location/Qualifiers  
 XX mat\_peptide 1675..4782  
 XX /tag- a  
 XX /product- "Haemophilus influenzae strain 12 mature HMW2A  
 protein"

XX WO200020609-A2.

XX 13-APR-2000.

XX 07-OCT-1999; 99WO-CA00938.

XX 07-OCT-1998; 98US-0167568.

XX 08-DEC-1998; 98US-0206942.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Yang Y, Klein MH;

XX P-PSDB; AAB01849.

XX Nucleic acid molecule for producing recombinant high molecular weight  
 XX proteins of Haemophilus which are used as a vaccine to provide  
 XX protection against Haemophilus induced diseases in humans -

XX Claim 7; Fig 29E-N; 307pp; English.

XX The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene  
 CC clusters termed hmwA1 and hmwA2. Each hmwABC operon comprises hmwA,  
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins  
 CC and the hmwB and hmwC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMW  
 CC proteins. The modified hmwABC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwA genes (AA52173-A52198)  
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae  
 CC strains Joyce, K1, K21, LCPD2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents a modified NTHi hmwA gene encoding a mature  
 CC HMWA protein.

XX Sequence 3263 BP; 1173 A; 628 C; 674 G; 788 T; 0 other;

XX Query Match 66.1%; Score 3263; DB 21; Length 3263;  
 XX Best Local Similarity 100.0%; Pred. No. 0;  
 XX Matches 3263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1675 CCTGATGATGTAACAAATTAAGACCGGAGCCCGCTTCGCAATATACCGGTATAATGAT 1734  
 DB 1 CCTGATGATGTAACAAATTAAGACCGGAGCCCGCTTCGCAATATACCGGTATAATGAT 60

QY 1735 GAATTCACAGCAGCGGTGAAGCAAGGACCCCTAAATAAATAATAGCGAACTCAAAACA 1794  
 DB 61 GAATTCACAGCAGCGGTGAAGCAAGGACCCCTAAATAAATAATAGCGAACTCAAAACA 120  
 QY 1795 AGCTAACCAATACAACTATTTCAAATTTCTGAAAACGCGCTGGACAATGAATATAACG 1854  
 DB 121 AGCTAACCAATACAACTATTTCAAATTTCTGAAAACGCGCTGGACAATGAATATAACG 180  
 QY 1855 GCATCAAGAAAACCTTACCGTTAATAGCTCAATCAATCAATCGGAAGCACTCCCACTTAAT 1914  
 DB 181 GCATCAAGAAAACCTTACCGTTAATAGCTCAATCAATCAATCGGAAGCACTCCCACTTAAT 240  
 QY 1915 CTCCATAGTAAGGTCAGCGTGGCGGAGCGGCTTCAGATTGATGGAGATATTACTTTCTAAA 1974  
 DB 241 CTCCATAGTAAGGTCAGCGTGGCGGAGCGGCTTCAGATTGATGGAGATATTACTTTCTAAA 300  
 QY 1975 GCGGGAATTTAAACCATTTTATCTGGCGGATGGGTTGATGTTCTATATAAATAATATTACGCTT 2034  
 DB 301 GCGGGAATTTAAACCATTTTATCTGGCGGATGGGTTGATGTTCTATATAAATAATATTACGCTT 360  
 QY 2035 GATCAGGGTTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAAATAACAAA 2094  
 DB 361 GATCAGGGTTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAAATAACAAA 420  
 QY 2095 GCACGCGAGCGGCAAAATGCTTAAATTTGTCGCCAGGCGACCTGTAACCATTTACAGGAGAG 2154  
 DB 421 GCACGCGAGCGGCAAAATGCTTAAATTTGTCGCCAGGCGACCTGTAACCATTTACAGGAGAG 480  
 QY 2155 GGAAGAATTTTCAGGCTTAAACAGCTTCTTTAAACGGAACGCGTAAAGGCTGGAATATC 2214  
 DB 481 GGAAGAATTTTCAGGCTTAAACAGCTTCTTTAAACGGAACGCGTAAAGGCTGGAATATC 540  
 QY 2215 ATTTCATCAGTGAATTAATTTAAACCAATCTTAGTGCGACAAATTAACATATCTGGGAT 2274  
 DB 541 ATTTCATCAGTGAATTAATTTAAACCAATCTTAGTGCGACAAATTAACATATCTGGGAT 600  
 QY 2275 ATAACAAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAAC 2334  
 DB 601 ATAACAAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAAC 660  
 QY 2335 CACTGGAACGCTCAGTCTCTTAACTTAGAGACAGGCGCAAAATTTTAAATATAC 2394  
 DB 661 CACTGGAACGCTCAGTCTCTTAACTTAGAGACAGGCGCAAAATTTTAAATATAC 720  
 QY 2395 ATTTCAGCAATAGCAAAAGGCTTAAACCAACACAGTATAGAAGCTCTCGAGGGGTGAATTT 2454  
 DB 721 ATTTCAGCAATAGCAAAAGGCTTAAACCAACACAGTATAGAAGCTCTCGAGGGGTGAATTT 780  
 QY 2455 AACGGGTAAATGCAACATGTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 2514  
 DB 781 AACGGGTAAATGCAACATGTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 840  
 QY 2515 TTAACCAACAGGAGACATGAACCAACCAACCTTACCAATTCGGTTTTTATAGCAAT 2574  
 DB 841 TTAACCAACAGGAGACATGAACCAACCAACCTTACCAATTCGGTTTTTATAGCAAT 900  
 QY 2575 ATCAGCCCACTGGTGGGCTCTGTTTTTTTATATATATATATATATATATATATATATATAT 2634  
 DB 901 ATCAGCCCACTGGTGGGCTCTGTTTTTTTATATATATATATATATATATATATATATATAT 960  
 QY 2635 GGGGCTGAGTTAAATAATGAGTGAATTAATATCTCTAACCGCGCTTAATTTTCTTAAAT 2694  
 DB 961 GGGGCTGAGTTAAATAATGAGTGAATTAATATCTCTAACCGCGCTTAATTTTCTTAAAT 1020  
 QY 2695 TCCCATGTTGCGGCGGATGACGCTTTTAAATCAACAAAGACTTAAACCATTAATATGCAAC 2754  
 DB 1021 TCCCATGTTGCGGCGGATGACGCTTTTAAATCAACAAAGACTTAAACCATTAATATGCAAC 1080  
 QY 2755 AATTCAATTTACGCTTCAGACAGACGAAAGATGATTTTTTATGACGGGTACGCGCAAT 2814  
 DB 1081 AATTCAATTTACGCTTCAGACAGACGAAAGATGATTTTTTATGACGGGTACGCGCAAT 1140  
 QY 2815 GCCATCAATTCACACTACACATATCCATTTCTGGCGGCTAATGTCACCCCTTGGTGACAA 2874

Db 1141 GCCATCAATTCACACTACACATATCCATCTCGGGGTAATGTCAACCCCTTGGTGACAA 1200  
QY 2875 AACTCAAGCAGCAGCATTACGGGAATATTTACTATCGAGAAAGCACAATGTTCAGCTA 2934  
Db 1201 AACTCAAGCAGCAGCATTACGGGAATATTTACTATCGAGAAAGCACAATGTTCAGCTA 1260  
QY 2935 GAAGCCAAATACGCCCTTAATCAGCAAAACATAAAGGATAGAGTTATAAACTTGGCAGC 2994  
Db 1261 GAAGCCAAATACGCCCTTAATCAGCAAAACATAAAGGATAGAGTTATAAACTTGGCAGC 1320  
QY 2995 TTGCTGCTTAATGGGAGTTTAACTTTCAGCGAAATGCGAGATATTAAGGCAATCTC 3054  
Db 1321 TTGCTGCTTAATGGGAGTTTAACTTTCAGCGAAATGCGAGATATTAAGGCAATCTC 1380  
QY 3055 ACTATTTACAGAGCGCCACTTTTAAAGGAAAGACTTAGAGATACCTTAATATCACCGGC 3114  
Db 1381 ACTATTTACAGAGCGCCACTTTTAAAGGAAAGACTTAGAGATACCTTAATATCACCGGC 1440  
QY 3115 AATTTTACCNAATANTGGCACTGCCGAAATTAATATTAACACAAGGAGTGGTAAACTTGGC 3174  
Db 1441 AATTTTACCNAATANTGGCACTGCCGAAATTAATATTAACACAAGGAGTGGTAAACTTGGC 1500  
QY 3175 AATGTTTACCAATGATGGTGATTTAAACATTTACCACTTAAAGGAAAGGAGTGGTAAACTTGGC 3234  
Db 1501 AATGTTTACCAATGATGGTGATTTAAACATTTACCACTTAAAGGAAAGGAGTGGTAAACTTGGC 1560  
QY 3235 ATCATCGGCGGAGATATATCAACAAAAGGAAAGCTTAAATATTTACAGACAGTAATAT 3294  
Db 1561 ATCATCGGCGGAGATATATCAACAAAAGGAAAGCTTAAATATTTACAGACAGTAATAT 1620  
QY 3295 GATGCTGAAATCCAAATTTGGCGCAATATCTCGCAAAAGGAAAGGAGTGGTAAACTTGGC 3354  
Db 1621 GATGCTGAAATCCAAATTTGGCGCAATATCTCGCAAAAGGAAAGGAGTGGTAAACTTGGC 1680  
QY 3355 TCCGATATAATTAATATCAACAAAAGGAAAGCTTAAATATTTACAGACAGTAATAT 3414  
Db 1681 TCCGATATAATTAATATCAACAAAAGGAAAGCTTAAATATTTACAGACAGTAATAT 1740  
QY 3415 TCTAGTTCAGATGCGACAGTAATGCCAACCTTAATATTTAAACCAAGAAATTTGAAATTTG 3474  
Db 1741 TCTAGTTCAGATGCGACAGTAATGCCAACCTTAATATTTAAACCAAGAAATTTGAAATTTG 1800  
QY 3475 ACAGAGACCTTAAGTATTTACAGTTCATTAAGCAGAGATTTACAGCAAAAGATGGTAGA 3534  
Db 1801 ACAGAGACCTTAAGTATTTACAGTTCATTAAGCAGAGATTTACAGCAAAAGATGGTAGA 1860  
QY 3535 GATTTAACTAATGCGACAGTAATGACGGTAAACAGCGGTGCCGAGCCAAACAGTAAC 3594  
Db 1861 GATTTAACTAATGCGACAGTAATGACGGTAAACAGCGGTGCCGAGCCAAACAGTAAC 1920  
QY 3595 TTTAACTAATGTTAAAGATTTCAAAATCTCTGCTGACGGTCAATGTGACACTAAATAGC 3654  
Db 1921 TTTAACTAATGTTAAAGATTTCAAAATCTCTGCTGACGGTCAATGTGACACTAAATAGC 1980  
QY 3655 AAAGTGAACACATCTAGAGCAATGCGGAGCTGAAAGCAATAGCGACACAGTACCGGC 3714  
Db 1981 AAAGTGAACACATCTAGAGCAATGCGGAGCTGAAAGCAATAGCGACACAGTACCGGC 2040  
QY 3715 TTAATTTACTGCAAAATGTAAGATTAACAAAAGATATTTACTTCTCAAAACAGTA 3774  
Db 2041 TTAATTTACTGCAAAATGTAAGATTAACAAAAGATATTTACTTCTCAAAACAGTA 2100  
QY 3775 AATATCCCGCTGGGAAAGGTTTACCACACAGCAGGCTTCGACCTTAAACGCAACAAAT 3834  
Db 2101 AATATCCCGCTGGGAAAGGTTTACCACACAGCAGGCTTCGACCTTAAACGCAACAAAT 2160  
QY 3835 GGCAACCAAGTATTACAAACAAACAGGTGATATCAGCGGTACGATTTCCGGTAAACAG 3894  
Db 2161 GGCAACCAAGTATTACAAACAAACAGGTGATATCAGCGGTACGATTTCCGGTAAACAG 2220  
QY 3895 GTAGTGTAGCGCGACTGGTGATTTAAACCACTAAATCCGGCTCAAAATTTAGAGCAAA 3954  
Db 3954 GTAGTGTAGCGCGACTGGTGATTTAAACCACTAAATCCGGCTCAAAATTTAGAGCAAA 4014

Db 2221 GTAAGTGTAGCGCGACTGGTGATTTAAACCACTAAATCCGGCTCAAAATTTAGAGCAAA 2280  
QY 3955 TCGGTGAGGCTTAATGTAACAGTGCACACAGTACAAATTTGGCGGTACAAATTTCCGGTAA 4014  
Db 2281 TCGGTGAGGCTTAATGTAACAGTGCACACAGTACAAATTTGGCGGTACAAATTTCCGGTAA 2340  
QY 4015 ACGGTAATGTTACGGCAAAAGCTGGCGATTTACAGTTGGGAATGGCGCAAGAAATTAAT 4074  
Db 2341 ACGGTAATGTTACGGCAAAAGCTGGCGATTTAAACAGTTGGGAATGGCGCAAGAAATTAAT 2400  
QY 4075 GCGACAGAGGAGCTGCAACCTTAAACCAACAGGGAATACCTTTGACTACTGAGAGCGGT 4134  
Db 2401 GCGACAGAGGAGCTGCAACCTTAAACCAACAGGGAATACCTTTGACTACTGAGAGCGGT 2460  
QY 4135 TCTAGCATCACTTCAACTAAGGGTACAGTACAGCTCTTGGCTCAGAAATGTTAGCATCGCA 4194  
Db 2461 TCTAGCATCACTTCAACTAAGGGTACAGTACAGCTCTTGGCTCAGAAATGTTAGCATCGCA 2520  
QY 4195 GGAAGCATTAATGCTGCTAATGTGACATTAATATCTACAGGCACTTTAAACCAAGGAGT 4254  
Db 2521 GGAAGCATTAATGCTGCTAATGTGACATTAATATCTACAGGCACTTTAAACCAAGGAGT 2580  
QY 4255 GGCTCGGATATTAAGCAACAGCGGCACTTGGTTATTAACGCAAAAGATGCTAAAGCTA 4314  
Db 2581 GGCTCGGATATTAAGCAACAGCGGCACTTGGTTATTAACGCAAAAGATGCTAAAGCTA 2640  
QY 4315 AATGCTGATCATCAGGTGATAGTACAGAGTGAATGCACTGGGATTTAAACACAGTAAT 4374  
Db 2641 AATGCTGATCATCAGGTGATAGTACAGAGTGAATGCACTGGGATTTAAACACAGTAAT 2700  
QY 4375 AGTGTGACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAT 4434  
Db 2701 AGTGTGACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAT 2760  
QY 4435 GGGTTAAATATCACTTCGAAGATGTTAGAGAACACTGTGGCTTAAAGAGCAAGAAAT 4494  
Db 2761 GGGTTAAATATCACTTCGAAGATGTTAGAGAACACTGTGGCTTAAAGAGCAAGAAAT 2820  
QY 4495 GAGTGAATATATCCAGCGAGGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAG 4554  
Db 2821 GAGTGAATATATCCAGCGAGGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAGCAAGTGTAG 2880  
QY 4555 GTCTTGAAGAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAACTTGGT 4614  
Db 2881 GTCTTGAAGAAAGTAAAGATTTATCTGATGAAGAAAGAAACATTTAGCTAACTTGGT 2940  
QY 4615 GTAAAGTGTAGCTTGTGTTAGCCAAATTAATACAAATTAACAGTCAATACACAAATGAA 4674  
Db 2941 GTAAAGTGTAGCTTGTGTTAGCCAAATTAATACAAATTAACAGTCAATACACAAATGAA 3000  
QY 4675 TTTAAACAGAGCCGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 4734  
Db 3001 TTTAAACAGAGCCGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 3060  
QY 4735 GGTAAAGCGCAGGAGTATGTAACCAATGTTGCTGAGGATGGAACCGGTAGTCAAGTAA 4794  
Db 3061 GGTAAAGCGCAGGAGTATGTAACCAATGTTGCTGAGGATGGAACCGGTAGTCAAGTAA 3120  
QY 4795 GACAAGTGAATTTCACTCTGCAATGAAGTCAATTTATTTTCTGATTTATTTACTGTGG 4854  
Db 3121 GACAAGTGAATTTCACTCTGCAATGAAGTCAATTTATTTTCTGATTTATTTACTGTGG 3180  
QY 4855 GTTAAAGTTCAGTACGGGCTTTACCCATCTTGTAAAAAATTTACGGAGAAATACAAATAG 4914  
Db 3181 GTTAAAGTTCAGTACGGGCTTTACCCATCTTGTAAAAAATTTACGGAGAAATACAAATAG 3240  
QY 4915 ATTTTAAACAGGTTATTTATG 4937  
Db 3241 ATTTTAAACAGGTTATTTATG 3263

RESULT 8  
AAQ49506

ID AAQ49506 standard; DNA; 5116 BP.  
 AC AAQ49506;  
 XX 26-APR-1994 (first entry)  
 DT  
 XX Sequence encoding high molecular weight protein 1 (HWM1).  
 DE  
 XX  
 XX  
 XX  
 KW HWM; high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae; ss.  
 XX  
 XX Haemophilus influenzae.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 CDS 351..4961  
 FT /\*tag= a  
 FT /product= high molecular weight protein 1  
 FT misc\_difference 360..362  
 FT /\*tag= b  
 FT /transl\_except= CTA encodes Isoleucine.  
 FT misc\_difference 642..644  
 FT /\*tag= c  
 FT /transl\_except= GAU encodes Alanine.  
 FT misc\_difference 1437..1439  
 FT /\*tag= d  
 FT /transl\_except= AAG encodes Asparagine.  
 FT misc\_difference 1674..1676  
 FT /\*tag= e  
 FT /transl\_except= CCG encodes Phenylalanine.  
 XX  
 XX W09319090-A.  
 PN  
 XX  
 XX  
 XX 30-SEP-1993.  
 PD  
 XX  
 XX 16-MAR-1993; 93WO-US02166.  
 PF  
 XX  
 XX 16-MAR-1992; 92GB-0005704.  
 PR  
 XX  
 XX (BARE/) BARENKAMP S J  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA  
 XX  
 XX Barenkamp SJ;  
 PI  
 XX  
 XX WPI; 1993-320683/40.  
 DR P-PSDB; AAR41723.  
 XX  
 XX High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 XX  
 XX Claim 3; Figure 1; 100pp; English.  
 PS  
 XX  
 XX The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 XX  
 XX  
 XX Sequence 5116 BP; 1766 A; 1008 C; 1037 G; 1305 T; 0 other;  
 SQ  
 Query Match 54.3%; Score 2680.2; DB 14; Length 5116;  
 Best Local Similarity 73.5%; Pred. No. 0;  
 Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps 16;  
 Qy 41 ATGCAACACACATTTACACACCTTTTTCAGCTCTATATGCATATTTTAAATAAT 100  
 Db 41 ATGCAACACACATTTACACACCTTTTTCAGCTCTATATGCATATTTTAAATAAT 99  
 Qy 101 AGTATAATCCGCCATATAAATGGTATATCTTTTCATCTTTTAACTTTTCATC 160  
 Db 100 AGTATAATCCGCCATATAAATGGTATATCTTTTCATCTTTTAACTTTTCATC 159  
 Qy 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220  
 Db 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC

Db 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219  
 Qy 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 280  
 Db 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 279  
 Qy 281 GAATGAAGAGGAGCTGAACGACGCAATTAAGTAATTTAAATTTGTTCACTAACCT 340  
 Db 280 GAATGAAGAGGAGCTGAACGACGCAATTAAGTAATTTAAATTTGTTCACTAACCT 339  
 Qy 341 TAGGAGAAATATGAACAAGATATATCGTCTCAAAATTCAGCAACGCTGAAATCGTTGG 400  
 Db 340 TAGGAGAAATATGAACAAGATATATCGTCTCAAAATTCAGCAACGCTGAAATCGTTGG 399  
 Qy 401 TTGCTGTCTGAATTTGGCAGCGGGTTGTGACCATTTCCACAGAAAAGGCTTCCGCTATG 460  
 Db 400 TTGCTGTCTGAATTTGGCAGCGGGTTGTGACCATTTCCACAGAAAAGGCTTCCGCTATG 459  
 Qy 461 TTACTATCTTTAGGTGTAAACCACTTAGCGTTAAAGCCACTTTCCGCTATGTACTATCTT 520  
 Db 460 CTGCTCGCATGAAAGTGCGTCACTTAGCGTTAAAGCCACTTTCCGCTATGTACTATCTT 519  
 Qy 521 TAGGTGTAAACATCTATCCCAATCTGTTTTCAGCAAGCGCTTACAAGGAATGATGTAG 580  
 Db 520 TAGGTGTAAACATCTATCCCAATCTGTTTTCAGCAAGCGCTTACAAGGAATGATGTAG 579  
 Qy 581 TACACGGCAGACCCACTATGCAAGTAGATGTTAATAAACCATTTATCCGCAACAGTGTG 640  
 Db 580 TACACGGCAGACCCACTATGCAAGTAGATGTTAATAAACCATTTATCCGCAACAGTGTG 639  
 Qy 641 ACGCTATCATTAATTTGGAACAAATTTAAACATCGACCAAAATGAATGTGCGAGTTTTTAC 700  
 Db 640 ACGATATCATTAATTTGGAACAAATTTAAACATCGACCAAAATGAATGTGCGAGTTTTTAC 699  
 Qy 701 AAGAAACAAACAACTCCCGCTGTTTAAACCGGTTTACATCTAACCCAAATCTGTAACATAG 760  
 Db 700 AAGAAACAAACAACTCCCGCTGTTTAAACCGGTTTACATCTAACCCAAATCTGTAACATAG 759  
 Qy 761 AAGGATTTTAGATTTTAAACGAGCAAGTCTTTTAAACCGGTTTACATCTAACCCAAATCTGTAACATAG 820  
 Db 760 AAGGATTTTAGATTTTAAACGAGCAAGTCTTTTAAACCGGTTTACATCTAACCCAAATCTGTAACATAG 819  
 Qy 821 GTAAAGACGCAATTTAAACATCTTACCGCTTTACGCTTACGCTAGACATTTCTAACG 880  
 Db 820 GTAAAGACGCAATTTAAACATCTTACCGCTTTACGCTTACGCTAGACATTTCTAACG 879  
 Qy 881 AAAACATCAAGCGCGGTATTTTACCTTCGAGCAACCAAGATTAAGCGCTCGCTGAAA 940  
 Db 880 AAAACATCAAGCGCGGTATTTTACCTTCGAGCAACCAAGATTAAGCGCTCGCTGAAA 939  
 Qy 941 TTGTGAATCAGCGTTTAAATTTACTGCTGTAAGAGCGGAGTGAATCTTATTGTTGCA 1000  
 Db 940 TTGTGAATCAGCGTTTAAATTTACTGCTGTAAGAGCGGAGTGAATCTTATTGTTGCA 999  
 Qy 1001 AAGTGAATAACGAGGTGTGATTAGCGTAAATGTGCGAGCATTTCTTTTACGCGAGGC 1060  
 Db 1000 AAGTGAATAACGAGGTGTGATTAGCGTAAATGTGCGAGCATTTCTTTTACGCGAGGC 1059  
 Qy 1061 AAAAAATCACCATCAGCGATATATAAACCACCACTTACTTTACAGCATTTGCGCGCTG 1120  
 Db 1060 AAAAAATCACCATCAGCGATATATAAACCACCACTTACTTTACAGCATTTGCGCGCTG 1119  
 Qy 1121 AAAAAATCAGCGGTCAATCTGGCGATATTTTTCGCAAGCGGTAACTAATTAATGTCCTG 1180  
 Db 1120 AAAAAATCAGCGGTCAATCTGGCGATATTTTTCGCAAGCGGTAACTAATTAATGTCCTG 1179  
 Qy 1181 CTGCGCATTTTCGAACCAAGGTAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCG 1240  
 Db 1180 CTGCGCATTTTCGAACCAAGGTAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCG 1239  
 Qy 1241 GCAATATTTGTTCTTCCGCAAGAGGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTC 1300  
 Db 1240 GCAATATTTGTTCTTCCGCAAGAGGGTGAAGCGGAATTTGCGGTGTAATTTCCGCTC 1299

QY	1301	AAAAATCAGCAAGCTAAAGCGGGCAAGCTGATGATTACAGCGGATAAAGTCACATATAAAAA	1361
Db	1300	AA	1359
QY	1361	CAGTTCGAGTTATCGACCTTTTCAGGTAAAGAAGGGGAGAAACTTTACCTTTGGCGGTGACG	1420
Db	1360	CAGTTCGAGTTATCGACCTTTTCAGGTAAAGAAGGGGAGAAACTTTACCTTTGGCGGTGACG	1419
QY	1421	AGCGCGCGCAAGGTAAAAACGGCATTTCAATTTAGCAAGAAAAACCTCTTTAGAAAAAGCT	1480
Db	1420	AGCGCGCGCAAGGTAAAAAGGGCATTTCAATTTAGCAAGAAAAACCTCTTTAGAAAAAGCT	1479
QY	1481	CAACCATCAATGTATCAGGCAAAAGAAAAAGCGAGCGCCTATTGTGTGGGCGCATATTG	1540
Db	1480	CAACCATCAATGTATCAGGCAAAAGAAAAAGCGAGCGCCTATTGTGTGGGCGCATATTG	1539
QY	1541	CGTTTAATTCAGCGCAATTAACGCTCAAGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1600
Db	1540	CGTTTAATTCAGCGCAATTAACGCTCAAGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1599
QY	1601	TTGTGGAGACATCGGGGCATTTATTTATCCATTGACAGCAATGCAATTTGTTAAAAACAAAG	1660
Db	1600	TTGTGGAGACATCGGGGCATTTATTTATTCATCAAGACAAATGCAATTTGTGAGCGCAAG	1659
QY	1661	AGTGGTTGCTAGACCCCTGATGATGAAGTAAACAATTCAGCGCGAGAGACCCCTTCGCAATAATA	1720
Db	1660	AGTGGTTGTTAGACCCGGATATGTATCTTATATGCAAGAACAGCAGACGACGACAAATA	1719
QY	1721	CCGGTATAATGATGAATTTCCCAACGAGCACCGGTGAAGCAAGGACCTTAAAAAAATA	1780
Db	1720	CTTCAGAACAGCATGAATACACGGATCCCGGAATAGTGCCAGCACCCCAACAGCARACA	1779
QY	1781	GCAGACTCAAAACACCGTAAACCAATACAACTATTTCAAAATTCATGAAAAAGCGCTGGGA	1840
Db	1780	AAGAA---AAGCACCAATTAAACAACAACTCTTGAGAGTATACTAAAAAAGGTACCT	1836
QY	1841	CAATGAATATAACGGCATCAAGAAAACTTACCCTTTAATAGCTCAATCAACATCGGAAGCA	1900
Db	1837	TTGTTTAACATCACTGCTAATCAACGCATCTATGTCAATAGCTCCATTAATTT---ATTCCA	1893
QY	1901	ACTCCCACTTAATTTCTCCATAGTAAAGTCAGCGTGGGAGGCGGTTCAGATGTATGGAG	1960
Db	1894	ATGGCACTTAACCTCTTTGGAGTGAAGGTTCGAGCGGTGGCGCGGTGAGATTAACAACG	1953
QY	1961	ATATTAC-----TTCTAAAGCGGAATTTAACCATTTATCTGTGGCGGATCGG	2008
Db	1954	ATATTACCACCGGTGATACACAGAGTCAAACTTAAACATTTACTCAGCGGGCTGGG	2013
QY	2009	TTGATGTTCAATAAAATATTAGCTTTGATCAGGGTTTTTAAATATTATCCGCGCGCTCCG	2068
Db	2014	TTGATGTTCAATAAAATATCTCACTCGGGCGCAAGGTAAACATAACATTCACGCTAAC	2073
QY	2069	TAGCTTTTGAAGGTGGAATATAACAAGCAGCGACCGCGCAAAATGCTAAATTTGTCCGCC	2128
Db	2074	AAGATATCGCTTTGAGAAAGGAGCAACCAAGTC-----ATTACAGGTC	2118
QY	2129	AGGCACCTGTAAACCATTTACAGGAGAGGAAAAGATTTCAGGGCTTAACAACGPTCTTTAA	2188
Db	2119	AAGGGCATTTACC---TTCAGCAATCAAAAAGGTTTTAGATTTAATATGCTCTCTAA	2175
QY	2189	ACGGAACCGGTAAAGTCTGAATATATCTTTTCATCAGTCAATAA-----TTTAAACC	2239
Db	2176	ACGGCACTGGCAGCGGACGTGCAATTCACCACTTAAAGAACCAATAAATACGCTATCAAA	2235
QY	2240	ACAATCTTTAGTGGCAATTTAACATATATCTGGGAATAAACAAT---TAAACCAAACTACGA	2296
Db	2236	ATAAATTTGAAGGACCTTTTAAATATTTTCAGGGAAGTGAACATCTCAATGGTTTACCTTA	2295
QY	2297	GAAAGAACACCTCGTATGGCAAAACAGCCATGATTCGCGCTGGAACGTCAGTCTCTTA	2356
Db	2296	AAATGAAGTGGATGATATAAATTTCAAGAGCAGCACTTACTCGAATTTAACCTCTCTAA	2355

QY	2357	ATCT---	AGAGACAGCGCGCAATTTT	TAACCTTTT	TAATACATTT	CAAGCAANTAGCAAG	2413
Db	2356	ATGTTT	CGAGAGTGGCGAGTTT	AACCTC	TAATGACT	CCAGAGGAAGCGATAGCGAC	2415
QY	2414	GCTTAAC	ACACAGTATAGAAGCTCT	GCAAGGGTGAATTTT	AACGGCGTAAATGGCAAC	2473	
Db	2416	GCACACTT	ACCAGCCTTAT	ATTTAA	ACGGTATATCATTT	CAACAAGACACT--	2472
QY	2474	TGTCAT	TCANCTCA	AGAGGCGAAAGTTT	TAATTT	CAAAATTA	2533
Db	2473	TTAAT	GTGTGAACGAATGCAAGAGTCA	ACTTTG	CATCAAGGCACCAATAGGGAATAA	2532	
QY	2534	TGAAC	CACAAGCAACCTTT	ACCAATTC	CGTTTT	AGCCAAATATCACAGCCACTGTGGGG	2593
Db	2533	AGTAT	CTAGTTTGA	ATTAACGCATCAT	TAAT	TGGAACAATTTCACTTT	2592
QY	2594	GCTCTG	TTTT	TTTTCATATATATG	CCAACCACTTCTG	CAGAGGGGCTGAGTTAA	2647
Db	2593	GTGTTG	ATTTCACTTCTG	CCCTCATCTT	AGGTC	CAACCCCGGTGTAGTTATAA	2652
QY	2648	AAATG	AGTGAATTAATATCT	TAACGGCGCTAATTT	TACCTTT	AAATTTCCCATGTTCCGG	2707
Db	2653	ATTCT	AAATACTTTT	ATGTTTCA	CAGGCTCAAGTTT	TAGATTTTAA	2712
QY	2708	CGATG	ACGCTTTT	AAAAATCA	CAAGAGCTTAAC	CAATAAATGCAACCAATTTCAATTTCA	2767
Db	2713	CAAAA	ACTGGCTCTCA	ATAGAGANAATTT	TAAC	TTTAATGCCACC-----	2766
QY	2768	GCCTC	ACAGAGAAAGATGATTTT	TATG	ACGGGTAC	CGCAGCAGNATGCCATCAATCA	2827
Db	2767	ACATA	CACTTTTTC	CAAGTTGA	AGCACCAGTGAATGATTT	GGTAAAGCATTTGATAGCA	2826
QY	2828	CCTACA	ATATCAATTTCTG	CGCGGTATG	TAC	CCCTTGGTGACAAACTCAAGCAGCA	2887
Db	2827	AAAAA	ACATAACCTTTG	AGGAGGTAC	ATAC	CTACCTTGGCTCCAGGAAGCGTTAACG	2886
QY	2888	GCAT	TACGGGATATTTACT	ATCGAAGAGCAGCAAT	GTATAC	GCTAGAACCAATAACG	2947
Db	2887	AAAT	CGAAGGCAATGTACT	ATCAAT	TAACAACGCTAAC	CTCACTCTTATCGGTTCCGGATT	2946
QY	2948	CCCCT	ATATCAGCAAAACAT	AAGGATAG	AGTTAT	AAAACTTGGCAGCTTCTCGCTTAATG	3007
Db	2947	TTGAC	ACCATCAAAA-----	ACCTTT	TA	CTTTAAAAAGATGTCAATCAATAATAGC	3000
QY	3008	GGAG	TTTAAGTTTAACTGG	CGAAAATG	CAGATATTA	AAAGCAATCTCACTATTTACAGAA	3067
Db	3001	GCAAC	CTTACCGCTGAGGCAAT	ATTGT	CAATATAC	CGGAAATCTTACCGTTGAAGTA	3060
QY	3068	GGCG	CACTTTTAAGGA	AAAGACT	TAGAGATAC	CCCTAAATATCACCGCAATTTTACCAATA	3127
Db	3061	ACG	CTAATTTCAAGCTAT	CACAAATTT	CACTTTT	TAATGATAGGCGGCTTTGTTGACACA	3120
QY	3128	ATGG	CACTCGCGAAATTAAT	ATAACA	CAGAGGTGT	AAACTTGGCAATGTTACCAATG	3187
Db	3121	AAGC	CAATTC	AAATATTTCC	ATTGCC	AAAGGGGCTCGCTTTAAGACATGTAATTT	3180
QY	3188	ATGG	TGATTTAA	CAATTAAC	CTACGCT	TAACGCAACCAAGACATCATCGCGGAG	3247
Db	3181	CCA	AGATTTTAAGCATC	CACCAAACTCC	AGCTT	TACGCACTTATTAAGCGCA	3240
QY	3248	ATAT	ATCAACAAAAAG	GAAGCTTAA	ATATAC	AGCAGTAAATGATGCTGAAATCC	3307





XX

PT New immunogenic high mol. wt. proteins of non typeable

XX Hemophilus - useful in protective vaccines

XX

PS Claim 2; Page 28; 127pp; English.

XX

CC The HmW1 protein encoded by this sequence is useful in a vaccine to  
 CC protect against disease caused by non-typeable Haemophilus which are  
 CC not controlled by H. influenzae type b (Hib) vaccines. The encoded  
 CC protein can also be used as a carrier for protective Hib  
 CC polysaccharide (in a conjugate vaccine against meningitis) or for  
 CC other antigens, haptens, etc. This DNA sequence is part of a larger  
 CC sequence encoding a gene cluster for the hmw1 gene (9011 bp).

XX

SQ Sequence 5116 BP; 1766 A; 1008 C; 1037 G; 1305 T; 0 other;

Query Match

Best Local Similarity 54.3%; Score 2680.2; DB 15; Length 5116;

Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps

16;

QY 41 ATGACAAACAACAAATTTACAAACACCTTTTTCGAGCTATATATGCAAAATATTTAAAAAAT 100

DB 41 ATGACAAACAACAAATTTACAAACACCTTTTTCGAGCTATATATGCAAAATATTTAAAAAAT 99

QY 101 AGTATAAATCCGCCATATATAAATGGTATATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 160

DB 100 AGTATAAATCCGCCATATATAAATGGTATATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 159

QY 161 TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220

DB 160 TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219

QY 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 280

DB 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 279

QY 281 GAATGAAGAGGAGCTGAACGAACCAATGATATGAAGTAATTTAATTTGTTCAACTAACCT 340

DB 280 GAATGAAGAGGAGCTGAACGAACCAATGATATGAAGTAATTTAATTTGTTCAACTAACCT 339

QY 341 TAGGAGAAAATATGAACAGATATATCTCTCAAAATTCAGCAACCGCTGAATGCTTTGG 400

DB 340 TAGGAGAAAATATGAACAGATATATCTCTCAAAATTCAGCAACCGCTGAATGCTTTGG 399

QY 401 TTGCTGTCTCTGAATGGCAGCGGTTGTGACCATTCACAGAAAAGGCTTCGCGTATG 460

DB 400 TTGCTGTCTCTGAATGGCAGCGGTTGTGACCATTCACAGAAAAGGCTTCGCGTATG 459

QY 461 TTACTATCTTTAGGTGTAAACACCTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 520

DB 460 CTGCTCGCATGAAAGTGGCTCACCTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 519

QY 521 TAGGTGTAAACATCTATTCCCAATCTGTTTACGAGCGCTTACAGGAATGGATGTAG 580

DB 520 TAGGTGTAAACATCTATTCCCAATCTGTTTACGAGCGCTTACAGGAATGGATGTAG 579

QY 581 TACAGGCACAGCCACTATGCAAGTAGATGGTATATAAACCATTATCCGCAACAGTGTG 640

DB 580 TACAGGCACAGCCACTATGCAAGTAGATGGTATATAAACCATTATCCGCAACAGTGTG 639

QY 641 ACGTATCATTAATGGAACAATTTAACAATCCACCAAAATGAATGGTGGAGTTTAC 700

DB 640 ACGTATCATTAATGGAACAATTTAACAATCCACCAAAATGAATGGTGGAGTTTAC 699

QY 701 AAGAAAACAACACTCGCGCGTATTCAACCGTGTTCATCTAACCAAAATCTCCCAATTA 760

DB 700 AAGAAAACAACACTCGCGCGTATTCAACCGTGTTCATCTAACCAAAATCTCCCAATTA 759

QY 761 AAGGGATTTAGATTTCAACGGACAGTCTTTTAAATCAACCCAAATGGTATCACAAATAG 820

DB 760 AAGGGATTTAGATTTCAACGGACAGTCTTTTAAATCAACCCAAATGGTATCACAAATAG 819

QY 821 GTAAAGACGCAATTTATTAACTAATGGCTTTACGGCTTCTACGGCTAGACATTTCTTAACG 880

DB 820 GTAAAGACGCAATTTATTAACTAATGGCTTTACGGCTTCTACGGCTAGACATTTCTTAACG 879

QY 881 AAAACATCAAGCGCGCTAATTTCACTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAA 940

DB 880 AAAACATCAAGCGCGCTAATTTCACTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAA 939

QY 941 TTGTTCAATCAGCGTTTAAATTTACTGTCGTTAAGAGCGGAGTGTAAATCTTATTGGTGGCA 1000

DB 940 TTGTTCAATCAGCGTTTAAATTTACTGTCGTTAAGAGCGGAGTGTAAATCTTATTGGTGGCA 999

QY 1001 AAGTGAATAAAGAGGCTGTGATTTAGCGTAAATTTGGCGAGCATTTCTTTACTCCAGGGC 1060

DB 1000 AAGTGAATAAAGAGGCTGTGATTTAGCGTAAATTTGGCGAGCATTTCTTTACTCCAGGGC 1059

QY 1061 AAAAATCACCATCAGCGATATAATAAACCCCAACCATTTACTTTACAGCATTCGCCGCGCTG 1120

DB 1060 AAAAATCACCATCAGCGATATAATAAACCCCAACCATTTACTTTACAGCATTCGCCGCGCTG 1119

QY 1121 AAAATGAAGCGTCAATCTGGCGATATTTTGGCAAGCGGTTAACAATTAATGTCGGTG 1180

DB 1120 AAAATGAAGCGTCAATCTGGCGATATTTTGGCAAGCGGTTAACAATTAATGTCGGTG 1179

QY 1181 CTGCCACTATTTCGAAACCAAGTAAACTTTCTGCTGATTTCTGTAAGCAAAAGATAAAGCG 1240

DB 1180 CTGCCACTATTTCGAAACCAAGTAAACTTTCTGCTGATTTCTGTAAGCAAAAGATAAAGCG 1239

QY 1241 GCAATATTTGTTCTTTCCGCAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1300

DB 1240 GCAATATTTGTTCTTTCCGCAAGAGGCTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1299

QY 1301 AAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTTACAGCGCATAAAGTCACATTAATAA 1360

DB 1300 AAAATCAGCAAGCTAAAGCGGCAAGCTGATGATTTACAGCGCATAAAGTCACATTAATAA 1359

QY 1361 CAGGTGCGAGTTATCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1420

DB 1360 CAGGTGCGAGTTATCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1419

QY 1421 AGCGCGCGAAGGTAAAGCGGCAATTTCAATTAGCAAGAAACCTCTTTAGAAAGGCT 1480

DB 1420 AGCGCGCGAAGGTAAAGCGGCAATTTCAATTAGCAAGAAACCTCTTTAGAAAGGCT 1479

QY 1481 CAACCATCAATGTATCAGGCAAGAAAGAGCGGACCGCTATTGTTGGGCGGATATTG 1540

DB 1480 CAACCATCAATGTATCAGGCAAGAAAGAGCGGACCGCTATTGTTGGGCGGATATTG 1539

QY 1541 CGTTAATTCAGCGCAATATTAAAGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGTT 1600

DB 1540 CGTTAATTCAGCGCAATATTAAAGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGTT 1599

QY 1601 TTGTGGAGACATCGGGGCAATTTATTTATCCATTTGACAGCAATGCAATTTGTTAAACCAAG 1660

DB 1600 TTGTGGAGACATCGGGGCAATTTATTTATCCATTTGACAGCAATGCAATTTGTTAGCGCAAG 1659

QY 1661 AGTGGTTGCTAGACCTGTATGATGTAACAAATTTGAAGCGGAAGACCCCTTCGCAATAATA 1720

DB 1660 AGTGGTTGCTAGACCTGTATGATGTAACAAATTTGAAGCGGAAGACCCCTTCGCAATAATA 1719

QY 1721 CCGGTATAAATGATGAATTTCCCAACAGGCGGCTGAAGCAGCGCACCTTAACAAAATA 1780

DB 1720 CTTTCAAGACCATGATATACAGCGGATTCGGGAATAGTGGCAGCACCCCAACCAAGAA 1779

QY 1781 GCGAATCAAAACACGCTTAACCAATTAACAACTATTTCATAATTTATCTGAAACCGCTGGA 1840

DB 1780 AAGAA---AAGACAACATTAACAAACACAACTCTTTGAGAGTATCTAATAAAGAGTACCT 1836

QY 1841 CAATGAATATACGGCATCAGAAACAACTTACGTTAATAGCTCAATCAACATCGGAAGCA 1900

DB 1837 TTGTTAATCACTACTGCTAATCAACGCACTCTATGTTCAATAGCTTCCATTAATTT---ATCCA 1893

QY 1901 ACTCCACTTAATTTCTCCATAGTAAAGGTACGCTGGCGGAGCGGCTTCAGATTGATGAG 1960

us-10-092-880-3.rng

Tue Mar 25 08:39:01 2003

Db 1894 ATGGCAGCTTAACCTTTTGGAGTGAAGGTCGGAGCGGTGGCGGCTTGAGATTAACAACG 1953  
Qy 1961 ATATTAC-----TTCTAAAGCGGAAATTTAAACCAATTTATCTGCGCGATGG 2008  
Db 1954 ATATTACCGCGGTGATGATACAGAGGTGCAAACTTTAACTTACTCAGCGCGCTGGG 2013  
Qy 2009 TTGATGTTTCAATAAATAATTTACCTTTGATCAGGGTTTTTAAATATTTACCGCGCTTCG 2068  
Db 2014 TTGATGTTTCAATAAATAATTTACTCTCGGCGCAAGGTAACTAAACATTTACAGCTAAAC 2073  
Qy 2069 TAGCTTTTGAAGTGAATAATAACAAGCACGCGACGCGCAATGCTAAATTTGTCCGCC 2128  
Db 2074 AGATATCGCTTTGAGAAAGGAACAACCGTC-----ATTACAGTTC 2118  
Qy 2129 AGGCACTGTAACTATACAGAGAGGGAAGATTTTCAGGGCTTAACAACGTATCTTTAA 2188  
Db 2119 AGGGACTATTAC---TCAGGCAATCAAAAGGTTTGAATTTAATGATGCTCTCTAA 2175  
Qy 2189 ACGGAACGGGTAAAGTCTGTAATATCATTTTCATCAGTGAATAA-----TTTAAACC 2239  
Db 2176 ACGGCACTGGCAGCGGACTGCAATTCACCACTAAAGAACCAATAAATAGCTATCAAA 2235  
Qy 2240 ACAATCTTAGTGGACAAATTAACATATCTGGGAATATAACAAT---TAACCAAACTACGA 2296  
Db 2236 ATAAATTTGAAGGACTTTAAATATTTTCAGGGAAGTGAACATCTCAATGGTTTTACCTA 2295  
Qy 2297 GAAAGAACACCTCGTATTGGCAACACAGCCATGATTCGCACTGGAAGTCAAGTCTCTTA 2356  
Db 2296 AAAATGAAAGTGGATATAATTAATTCAGAGGAGCGCACTTACTGGAATTTAACTCTCTAA 2355  
Qy 2357 ATCT---AGACAGCGCGCAAAATTTTACCTTTATTAATATACATTTCAAGCAATAGCAAG 2413  
Db 2356 ATGTTTCGAGAGTGGCGAGTTTAACTCACTATTGACTCCAGAGGAAGGATAGTGCAG 2415  
Qy 2414 GCTTAACAACACATGATGAAGCTCTGCGGGGTGAATTTTAAAGCGGTAAATTTAAAGCGCA 2473  
Db 2416 GCACACTTACCGCTTATTAATTTAAAGCGTATATCATTTCAACAAAGACACT---ACCT 2472  
Qy 2474 TGTCATTTCAATCTCAAGAGGAGCGAAGTTAATTTCAATTTAAACCAACAGAGAA 2533  
Db 2473 TTAATGTTGAAGCAATGCAAGAGTCAACTTTGACATCAAGGCACCAATAGGGATAAATA 2532  
Qy 2534 TGAACACAGCAACCTTTTACCAATTCGGTTTGGACCAATATCACGCCACTGCTGGG 2593  
Db 2533 AGTATCTAGTTTGAATTTACGATCAATTTAATGGAACATTTTCAAGTTTCGGAGGGGGA 2592  
Qy 2594 GCTCTGTTT-----TTTTTGATATATATGCCAACCTTCTGGCAGAGGGCTGAGTTAA 2647  
Db 2593 GTGTTGATTTACACTTCTCGCTCATCTCTAAGCTCCAAACCCCGGTGATGTTATA 2652  
Qy 2648 AAATGAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTAAATTTCCCATGTTCCG 2707  
Db 2653 ATTTCAATACTTTAATGTTTCAACAGGGTCAAGTTTAAAGATTTAAACTTTTCAAGCTCAA 2712  
Qy 2708 GCGATGACGCTTTTAAATCAACAAGACTTAAACCATAAATGCAACCAATTTCAATTTCA 2767  
Db 2713 CAAAAACTGGCTTCTCAATAGAGAAGATTTTAACTTTAAATGCCACC-----GGAGGCA 2766  
Qy 2768 GCTTCAGACAGACGAAAGATGATTTTATGCGGCTACGCAATGCCATCAATTTCAA 2827  
Db 2767 ACATAACACTTTTGAAGTTGAAGCCAGGATGGAATGATTTGGTAAAGGCATTTGAGCCA 2826  
Qy 2828 CCTACACATATCCATTTGCGCGCTAATGTCACCTTTGGTGGCAAAACATCAAGCACA 2887  
Db 2827 AAAAAACATAACCTTTGAAGGAGGTACATCACTTTGGCTCCAGGAAGCCGTAAACAG 2886  
Qy 2888 GCATTACGGGGAAATATTACTATCGAAGAACAGCAAAATTTAGCGTAGAACCAATACG 2947  
Db 2887 AAATCGAAGGCAATGTTACTATCAATAACAACGCTAACGCTCACTCTTATCGTTCCGAT 2946  
Qy 2948 CCCCTAATCAGCAAAACATAAGGGATPAGATTTATAAAAGTTGGCAGCTTGGCTGTTAATG 3007  
Db 2947 TTGACAACCATCAAAA-----ACCTTTAACTATTAAAAAAGATGTCATCAATTAATAGCG 3000

Qy 3008 GGAGTTTAAGTTTAACTGGCGAAATGCAGATATTAAGGCAATCTCAGTATTTTCAGAAA 3067  
Db 3001 GCACCTTTACCGCTGGAGCAATATTGTCAATATAGCGGAAATCTTACCGTTCAAAAGTA 3060  
Qy 3068 GCGGCACCTTTTAAAGGAAGACTAGAGATACCTTAATATCACCGCAATTTTACCATAA 3127  
Db 3061 ACGCTAAATTTCAAGCTATCACAAATTTTCACTTTTAAATGATAGGCGCTGTTTGACAACA 3120  
Qy 3128 ATGGCAGTCGCGAAATTAATATAACAAGAGTGGTAAAACTTTGGCAATCTTACCAATG 3187  
Db 3121 AAGGCAATTTCAAAATTTTCCATTTGCCAAAGAGGGGCTCGCTTTAAAGCAATTTAAT 3180  
Qy 3188 ATGGTGATTTAAACATTACCACTCACGCTAAACGCAACCAAGAGCATCATCGGCGAG 3247  
Db 3181 CCAAGAATTTTAAGCATCACCACTCCAGCTCCACCTTACCGCACTATTATAAGCGGCA 3240  
Qy 3248 ATATAATCAACAAAAAAGGAAGCTTTAAATATTACAGACAGTAATTAATGATGCTGAAATCC 3307  
Db 3241 ATATAACCAATAAAAACGGTGATTTAAATATTACGAAGAGGTAGTGATGATGAAATGC 3300  
Qy 3308 AAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTTCCGATAAAATTA 3367  
Db 3301 AAATTTGGCGCGCTGCTCGCAAAAGAGGTAATCTCACGATTTCTTCTGACAAAAATCA 3360  
Qy 3368 ATATCACCAACAGATTAACAATCAAAAGGATTTGATGAGAGGACTCTAGTTCCAGATG 3427  
Db 3361 ATATACCAACAGATTAACAATCAAGGAGGTGTTGATGCGGAGAAATCCGATTCAGAGC 3420  
Qy 3428 CGACAAGTAAATGCGCAACCTTAATTTAAACCAAGAAATTTGAATTTGACAGAGACCTAA 3487  
Db 3421 CGACAACCAATGCGCAATCTTAACCTTAACCAATTTAAACCAAGAAATTTTAAACGAACTTA 3480  
Qy 3488 GTATTTAGTGTTCATTAAGCAGATTTACAGCCAAAGATTTGATAGAGATTTTAACTATTG 3547  
Db 3481 ATATTTACAGTTTCAATTAAGCAGATTTACAGCTTAAGATGTTGATTTTAACTATTG 3540  
Qy 3548 GCAACAGTAAATGAGCGTAAACAGCGGTGCGGAAGCAATAGCGACACAGTAACTTTTAACTATTG 3607  
Db 3541 GTAACACCAATAGTGTCTGA---TGGTACTTAATGCCAAAGAAATTTTAAACGAGTTA 3597  
Qy 3608 AAGATTTCAAAATTTCTGCTGACGGTCAACAATGTCACACTTAAATAGCAAGTAAACAT 3667  
Db 3598 AAGATTTCAAAATTTCTGCTGACGGTCAACAAGGTGACACTTACACAGCAAGTGAACAT 3657  
Qy 3668 CTAGCAGCAATGCGGAGCTGAAAGCAATAGCGACACAGTACCGCTTAACTATTACTG 3727  
Db 3658 CCGTAGTAAATAACACACTGAAGATAGCATGACAATAATGCGGCTTAACTATCGATG 3717  
Qy 3728 CAAAAATGTAGAGTAAACAAAGATATTACTTCTTCAAAAAGTAAATATCACCGCGT 3787  
Db 3718 CAAAAATGTAAACAGTAAACAAATATTACTTCTCAAAAGCAGTAACTATCTCTCGCA 3777  
Qy 3788 C---GGAAAAGTTTACCCACAGCGCTCGACCTTAAAGCAACCAATTTGGA--- 3839  
Db 3778 CAAGTGAGAAATTTACCTTAAACAGGTACACCACTTACGCAACCACTGGTAACTGG 3837  
Qy 3840 ----- 3839  
Db 3838 AGATAACCGCTCAAAACAGGTAGTATCTAGTGGAAATTTAGTCCAGCTCTGGCTCTATAA 3897  
Qy 3840 ----- 3839  
Db 3898 CACTTACTGCAACCGAGGGCGCTCTTGTGTGTAGCAATATTTTCGGGCAACACCGTTACTG 3957  
Qy 3840 ----- 3839  
Db 3958 TTACTGCAATAGCGGTGCTAATAACACTTTGGCAGGCTCTTACAATTTAAAGAACCCGAGA 4017  
Qy 3840 -AGCAAGTATTACAAACCAAAACAGTATATCACGGTACGATTTCCGGTTAACACGCTAA 3898  
Db 4018 GTGTAAACCACTTCAAGTCAATCAGCGCATATCGGCGGTACGATTTCTTGGTGGCACAGTAG 4077

```
QY 3899 GTGTTAGCGCGACTGGTGATTTTAAACCACTAAATCCGGCTCAAAAATGAAGCGAAATCGG 3958
Db      ||||| ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4078 AGGTTAAAGCAACCGAAAGTTTAAACCACTAAATCCAAATTAAGCAACAAACAG 4137
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3959 GTGAGGCTAATGTAACAAGTGAACAGGTACAAATTCGGCGGTACAAATTCGGGTAAATACGG 4018
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4138 GCCAGGCTACGTAAACAGTGCNACAGGTACAAATTCGGGTACGATTCGGGTAAATACGG 4197
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4019 TAAATGTTACGGCAACCGTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATGCGA 4078
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4198 TAAATGTTACGGCAACCGTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATGCGA 4257
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4079 CAGAAGAGCTGCAACCTTAAACCGCAACAGGAATACCTTGACTACTGAAGCCGGTTCTA 4138
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4258 CAGAAGAGCTGCAACCTTAAACCGCAACAGGAATACCTTGACTACTGAAGCCGGTTCTA 4317
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4139 GCATCACTTCAACTAAGGCTAGGTAGACCTCTTGCTCAGAAATGGTACGATCGCAGGAA 4198
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4318 ACATTAATCTCAGCAAGGCTAGGTAAATCTTTCAGCTCAGATGGTAGCGTTCGAGGAA 4377
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4199 GCATTAATCTCAGTAAATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4258
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4378 GTATTAATCCCGCAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4437
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4259 CGGATATTAAGCAACCGTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATGCGA 4318
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4438 CAACATTAATGCAACCGTGGCGATTTAACAGTTGGGAATGGCGAGAAATTAATGCGA 4497
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4319 GTGATGTCATCAGTGTATGATACAGAAATGATGATGATGATGATGATGATGATGATGATG 4378
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4498 GCGCAGCATTTGGGTAAACCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4557
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4379 TGACTGCGGCAACCTCAAGCAGTGTGATATCACTGGGATTTAAACACAGTAATGGGT 4438
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4558 TAATGCGCAACCTCAAGCAGTGTGATATCACTGGGATTTAAACACAGTAATGGGT 4617
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4439 TAAATATCATTTCAAAAACCGTATAACACCGTACTGTTAAAGGCTTAAATGATGATG 4498
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4618 TGAATATCATTTCAAAAACCGTATAACACCGTACTGTTAAAGGCTTAAATGATGATG 4677
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4499 TGAATATCATTTCAAAAACCGTATAACACCGTACTGTTAAAGGCTTAAATGATGATG 4558
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4678 TGAATATCATTTCAAAAACCGTATAACACCGTACTGTTAAAGGCTTAAATGATGATG 4737
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4559 TTGAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAAATGATGATG 4618
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4738 TTGAAAAGTAAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAAATGATGATG 4797
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4619 GTGCTGTACGTTTGTGAGCCAAATTAATCAATTAATCAATTAATCAATTAATCAATTAAT 4678
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4798 GTGCTGTACGTTTGTGAGCCAAATTAATCAATTAATCAATTAATCAATTAATCAATTAAT 4857
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4679 CAACGACACCGTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTCTTCAAGTGGTA 4738
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4858 CAACGACACCGTCAAGTGAATTTCTGAAGTGAAGCGGTCTTCAAGTGGTA 4917
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4739 ATGCGGACGAGTATGACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACA 4798
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4918 ATGCGGACGAGTATGACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACA 4977
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4799 AGGTAGATTTTCATCTGCAATGAATCATTTTATTTCTGATTTATTTATTTATTTATTTAT 4858
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4978 AGGTAGATTTTCATCTGCAATGAATCATTTTATTTCTGATTTATTTATTTATTTATTTAT 5037
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4859 AAGTTACAGTACGGGCTTTACCCATCTTGTAAAAAATTAAGGAGATACAAATTAAGTATTT 4918
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5038 AAGTTACAGTACGGGCTTTACCCATCTTGTAAAAAATTAAGGAGATACAAATTAAGTATTT 5097
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4919 TTAACAGGTTATTTATG 4937
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5098 TTAACAGGTTATTTATG 5116
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 10  
AAT90994

ID AAT90994 standard; DNA; 5116 BP.

XX AC AAT90994;

XX DT 14-APR-1998 (first entry)

XX DE Non-typeable Haemophilus high mol.wt. surface protein hmw1 gene.

XX KW Non-typeable Haemophilus; high molecular weight surface protein; hmw1; hmw1 gene; immunogen; vaccine; otitis media; ss.

XX OS Haemophilus influenzae strain 12.

XX PH Key Location/Qualifiers

FT RBS 341..345

FT repeat\_unit /\*tag= a

FT repeat\_unit /\*tag= b

FT repeat\_unit /\*tag= c

FT repeat\_unit /\*tag= d

FT repeat\_unit /\*tag= e

FT repeat\_unit /\*tag= f

FT repeat\_unit /\*tag= g

FT repeat\_unit /\*tag= h

FT repeat\_unit /\*tag= i

FT repeat\_unit /\*tag= j

FT repeat\_unit /\*tag= k

FT repeat\_unit /\*tag= l

FT repeat\_unit /\*tag= m

FT repeat\_unit /\*tag= n

FT repeat\_unit /\*tag= o

FT repeat\_unit /\*tag= p

FT repeat\_unit /\*tag= q

FT repeat\_unit /\*tag= r

FT repeat\_unit /\*tag= s

FT repeat\_unit /\*tag= t

FT repeat\_unit /\*tag= u

FT repeat\_unit /\*tag= v

FT repeat\_unit /\*tag= w

FT repeat\_unit /\*tag= x

FT repeat\_unit /\*tag= y

FT repeat\_unit /\*tag= z

FT repeat\_unit /\*tag= aa

FT repeat\_unit /\*tag= ab

FT repeat\_unit /\*tag= ac

FT repeat\_unit /\*tag= ad

FT repeat\_unit /\*tag= ae

FT repeat\_unit /\*tag= af

FT repeat\_unit /\*tag= ag

FT repeat\_unit /\*tag= ah

FT repeat\_unit /\*tag= ai

FT repeat\_unit /\*tag= aj

FT repeat\_unit /\*tag= ak

FT repeat\_unit /\*tag= al

FT repeat\_unit /\*tag= am

FT repeat\_unit /\*tag= an

FT repeat\_unit /\*tag= ao

```

XX      /rpt_type= INVERTED
XX      W09736914-AL.
XX      09-OCT-1997.
XX      01-APR-1997; 97WO-US04707.
XX      01-APR-1996; 96US-0617697.
XX      (BARE/) BARENKAMP S J.
XX      Barenkamp SJ;
XX      WPI: 1997-503038/46.
XX      P-PSDB; AAW30293.
XX      High molecular weight proteins of non-typeable Haemophilus
XX      influenzae - useful for vaccine production
XX      Claim 6; Page 63-66; 183pp; English.
XX      This nucleic acid comprises the hmw1 gene of non-typeable
XX      Haemophilus influenzae strain 12 that encodes high molecular
XX      weight surface protein HMW1 (see AAW30293). A phage genomic
XX      library of strain 12 was screened for clones expressing high
XX      mol.wt. proteins using a high titre antiserum against HMWs.
XX      Strongly reactive clones were subcloned into T7 expression
XX      plasmid; all expressed either 125 kDa HMW1 or 120 kDa HMW2 (see
XX      AAW30294). The expressed proteins are truncated, starting at
XX      residue 442 of both full-length HMW1 and HMW2 gene products.
XX      Correct processing requires the products of additional
XX      downstream genes (see AAT90996 and AAT90997). Nucleotide sequences
XX      (see AAT90992-93) encoding 2 HMW proteins of non-typeable H.
XX      influenzae strain 5 (see AAW30291-92) have also been identified.
XX      The HMW proteins, conjugates and peptides can be used in
XX      vaccines, as immunogens for preparation of antibodies and as
XX      antigens for detection of these antibodies. The nucleic acid
XX      sequences can be used as to prepare recombinant proteins and as
XX      probes for detection of related genes.
XX      Sequence 5116 BP; 1766 A; 1009 C; 1036 G; 1305 T; 0 other;
XX
XX      Query Match      54.3%; Score 2680.2; DB 18; Length 5116;
XX      Best Local Similarity 73.5%; Pred. No. 0;
XX      Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps
XX      16;
XX
XX      QY 41 ATGACAAACACAAATTAGAACACACCTTTTTCAGTCTATATGCAAAATATTTTAAAAAAT 100
XX      DB 41 ATGACAAACACAAATTAGAACACACCTTTTTCAGTCTATATGCAAAATATTTT-AAAAAAT 99
XX
XX      QY 101 AGTATAAATCCGCCATATAAAATGGTAAATCTTTTCATCTTTTCATCTTTTTCATCTTCATC 160
XX      DB 100 AGTATAAATCCGCCATATAAAATGGTAAATCTTTTCATCTTTTCATCTTTTTCATCTTCATC 159
XX
XX      QY 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220
XX      DB 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219
XX
XX      QY 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 280
XX      DB 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 279
XX
XX      QY 281 GAATGAGAGGAGCTGAGCAACCAATGATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 340
XX      DB 280 GAATGAGAGGAGCTGAGCAACCAATGATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 339
XX
XX      QY 341 TAGGAGAAAATATGACAAAGATATATCGTCTCAAAATTCAGCAAAAGCTGATCTTTGG 400
XX      DB 340 TAGGAGAAAATATGACAAAGCTATATCGTCTCAAAATTCAGCAAAAGCTGATCTTTGG 399
XX
XX      QY 401 TTGCTGTGTCTGAATTTGGCAGCGGGTGTGTGACCATTTCCACAGAAAAGCTTCCGCTATG 460
XX      DB 401 TTGCTGTGTCTGAATTTGGCAGCGGGTGTGTGACCATTTCCACAGAAAAGCTTCCGCTATG 459

```

```

DB 400 TTGCTGTGTCTGAATTTGGCAGCGGGTGTGTGACCATTTCCACAGAAAAGCTGACGAAAAAC 459
QY 461 TTACTATCTTTAGGTGTAAACCACTTAGCGTTAAAGCCCACTTTCCGCTATGTTACTATCTT 520
DB 460 CTGCTCGCATGAAAGTGGGTCACTTAGCGTTAAAGCCCACTTTCCGCTATGTTACTATCTT 519
QY 521 TAGGTGTAAACATCTATTTCCACAACTGTTTGTAGCAAGCGGCTTACAAAGAAATGGATGTAG 580
DB 520 TAGGTGTAAACATCTATTTCCACAACTGTTTGTAGCAAGCGGCTTACAAAGAAATGGATGTAG 579
QY 581 TACAGGGCACGCCACTATGCAAGTAGTGTAAATAAACCATTATTCGCCAACACAGTGTG 640
DB 580 TACAGGGCACGCCACTATGCAAGTAGTGTAAATAAACCATTATTCGCCAACACAGTGTG 639
QY 641 AGCTATCATTAATTTGGAAACCAATTTAAACATCGACCAAAATGAAATGGTGTGAGTGTGTAC 700
DB 640 AGCATATCATTAATTTGGAAACCAATTTAAACATCGACCAAAATGAAATGGTGTGAGTGTGTAC 699
QY 701 AAGAAAACAAACAACTCCGCGGTATTCAACCGTGTATTCAACCGTGTATTCAACCGTGTATTCA 760
DB 700 AAGAAAACAAACAACTCCGCGGTATTCAACCGTGTATTCAACCGTGTATTCAACCGTGTATTCA 759
QY 761 AAGGGATTTTAGATCTTAACGGACAAAGTCTTTTAAATCAACCAAAATGATGTATCAACAATAG 820
DB 760 AAGGGATTTTAGATCTTAACGGACAAAGTCTTTTAAATCAACCAAAATGATGTATCAACAATAG 819
QY 821 GTAAAGACCAATTTATTAACACTTAATGGCTTTACGGCTTCTAGCTAGACATTTCTTAAGC 880
DB 820 GTAAAGACCAATTTATTAACACTTAATGGCTTTACGGCTTCTAGCTAGACATTTCTTAAGC 879
QY 881 AAAACATCAAGGCGCTAATTTACCTTCGACCAAAACCAAGATAAAGCGCTCGCTGAAA 940
DB 880 AAAACATCAAGGCGCTAATTTACCTTCGACCAAAACCAAGATAAAGCGCTCGCTGAAA 939
QY 941 TTCTGAATCACGGTTTAAATTTACTGTTCGTAAGACGCGAGTGTAAATCTTTATTGTGGCA 1000
DB 940 TTCTGAATCACGGTTTAAATTTACTGTTCGTAAGACGCGAGTGTAAATCTTTATTGTGGCA 999
QY 1001 AAGTGAACACGAGGTTGTATTAGCTAAATGTGTGCGAGCATTTCTTTACTGCGAGGCG 1060
DB 1000 AAGTGAACACGAGGTTGTATTAGCTAAATGTGTGCGAGCATTTCTTTACTGCGAGGCG 1059
QY 1061 AAAAATCACCATCAGCGATATATAAACCAACCACTTACTTACAGCATTTGCGCGCGCTG 1120
DB 1060 AAAAATCACCATCAGCGATATATAAACCAACCACTTACTTACAGCATTTGCGCGCGCTG 1119
QY 1121 AAAATGAAGCGGTCAATCTGGCGATATTTTTCGCAAGCGGTAAACATTAATGTCCGTG 1180
DB 1120 AAAATGAAGCGGTCAATCTGGCGATATTTTTCGCAAGCGGTAAACATTAATGTCCGTG 1179
QY 1181 CTGCCACTATTCGAAACCAAGGTAAACTTCTGCTGATCTGTAAAGCAAGATAAAGCG 1240
DB 1180 CTGCCACTATTCGAAACCAAGGTAAACTTCTGCTGATCTGTAAAGCAAGATAAAGCG 1239
QY 1241 GCAATATGTTCTTTCGCGCAAGAGGTTGAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1300
DB 1240 GCAATATGTTCTTTCGCGCAAGAGGTTGAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1299
QY 1301 AAAATCAGCAAGCTTAAAGCGCGCAAGCTGATGATTAACAGCGGATAAAGTCAQATTAATAA 1360
DB 1300 AAAATCAGCAAGCTTAAAGCGCGCAAGCTGATGATTAACAGCGGATAAAGTCAQATTAATAA 1359
QY 1361 CAGGTGCAAGTATTCGACCTTTTACAGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1420
DB 1360 CAGGTGCAAGTATTCGACCTTTTACAGTAAAGAGGGGAGAACTTACCTTTGGCGGTGACG 1419
QY 1421 AGCGCGCGAAGGTAAACCGCATTTCAATTTAGCAAGAAAACCTCTTTAGAAAAGGCT 1480
DB 1420 AGCGCGCGAAGGTAAACCGCATTTCAATTTAGCAAGAAAACCTCTTTAGAAAAGGCT 1479
QY 1481 CAACCATCAATGATCAGCAAGCAAGAAAAGCGGACCGCTATTGTGTGGGCGGATATTG 1540
DB 1480 CAACCATCAATGATCAGCAAGCAAGAAAAGCGGACCGCTATTGTGTGGGCGGATATTG 1539

```

Qy	1541	CGTTAAATTGACGCGCAATATTAAACGCTCAAGCGTAGTGCTGATATCGCTAAAAACCGGTGTT	1600
Db	1540	CGTTAAATTGACGCGCAATATTAAACGCTCAAGGTAGTGGTGATATCGCTAAAAACCGGTGTT	1599
Qy	1601	TTGTGGAGACATCGGGCATATTATTCATATGACAGCAATGCTTTAAAAACAAAG	1660
Db	1600	TTGTGGAGACGTCGGGCGATGATTATTCATCAAGACAAATGCAATTCCTGACGCCAAAG	1659
Qy	1661	AGTGGTTGCTAGACCCCTGATGATGAACAATTTGAAGCCGAGACGCCCTTCGCAATATA	1720
Db	1660	AGTGGTTGTTAGACCCGGATATGATATCTTAATGCAAGACAGCAGGACGACGACGAATA	1719
Qy	1721	CCGGTATAAATGATGAATTTCCCAACAGGCACCGGTGAAGCAAGCCACCTTAAAAAAATA	1780
Db	1720	CTTCAGAAGACGATGAATACACGGGATCGGGAAATAGTGCAGCACCCCAAAACGAAACA	1779
Qy	1781	GCGAACCTCAAAACACGCTTAACCATACAACTATTCAAAATATCTGAAAACGCCCTGGA	1840
Db	1780	AAGAA---AGACACACATTACAAACACAACTCTTTGAGAGTATACTAAAAAAGGTACCT	1836
Qy	1841	CAATGAATATAAGCGCATCAAGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCA	1900
Db	1837	TTGTTTAACATCACTGCTAATCAACGCATCTATGTCAATAGCTCCATTAATTT---ATCCA	1893
Qy	1901	ACTCCACACTAATTTCTCCATAGTAAAGGTCAGCGTGGCGGAGGCGTTCAGATGATGGAG	1960
Db	1894	ATGCGAGCTTAACCTCTTTCGAGTGAGGTCGGAGCGGTGGCGCGGTGAAGATTAACAACG	1953
Qy	1961	ATATTAC-----TTCATAAGCGGANAATTTAAACCATTTATCTTGGCGGATGGG	2008
Db	1954	ATATTACCACCGGTGATGATACAGAGGTGCAAACTTAACAATTTACTCAGCGGCTGGG	2013
Qy	2009	TTCATGTTCAATAAATATTACGCTTGATCAGGGTTTTTTAAATATTTACCGCGCTTCGG	2068
Db	2014	TTCATGTTCAATAAATATCTCACTCGGGCGCAAGGTAACATAAATTTACAGCTAAAC	2073
Qy	2069	TAGCTTTTGAAGGTGGAATATAACAAGCACGCGACGGCGAAATGCTAAAATTTGTCGCC	2128
Db	2074	AAGATATCGCTTTTGAGAAAGGAACCAACCAAGTC-----ATTACAGGTC	2118
Qy	2129	AGGCACTGTACCATTTACAGGAGAGGAAAAGATTTTCAGGGCTAACACAGCTATCTTTAA	2188
Db	2119	AAGGCACTATTACC---TCAGCAATCAAAAAGGTTTTAGATTTAATTAATGTCTCTCTAA	2175
Qy	2189	ACGGAACGGTTAAAGTCTGAATATCATTTTCATCAGTGAATA-----TTTAAACC	2239
Db	2176	ACGGCACTGGCAGCGACTGCAATTCACCACATAAAGAACCAATAATACGCTATCACAA	2235
Qy	2240	ACAATCTTAGTGGCACATTTACATATCTGGGAATATAACAAT---TAACCAAACTACGA	2296
Db	2236	ATAAATTTGAAGGGACTTTAAATATTTCAGGAAAGTGAACATCTCAATGGTTTTACCTA	2295
Qy	2297	GAAAGAACACCTCTGATTTGGCAAAACACGCCATGATTCGCACCTGGAAGCTCAGTGCTTA	2356
Db	2296	AAAATGAAAGTGTATGATATAATTAACAAGCAGCCACTACTGGAATTTAACTCTTAA	2355
Qy	2357	ATCT---AGACACAGGCCAAATTTTACCTTTATTAATACATTTCAAGCAATAGCAAAAG	2413
Db	2356	ATGTTTCGAGAGTGGCGAGTTTAACTCTACTATGACTCCAGAGGAAGCATATGTCAG	2415
Qy	2414	GCCTAACACACAGTATAGAAGCTCTCGACGGGTGAATTTTAACGGGTAAATGGCAACA	2473
Db	2416	GCACACTTACCAGCCTTATTAATTTAAACGGTATATCAATTCACAAGACACT---A	2472
Qy	2474	TGTCATTCAAATCTCAAGAAGGACGGAAGTTAATTTCAAAATTTAAAAACCAACGAGAACA	2533
Db	2473	TTAATGTTGAACGAAATGCAAGAGTCAACTTTTGACATCAAGGCCAATTAGGATATAATA	2532
Qy	2534	TGAACACACAGCAACCTTTACCAATTCGGTTTTAGCCAAATATCACAGCCACTGTGGGG	2593
Db	2533	AGTATCTTAGTTGAATTCAGCATCATTTAATCGAAACATTTTCAAGTTCTCGGAGGGGGA	2592

QY	2594	CTCTGTTT-----TTTTGATATATATAGCCACCACTTCTGGCAGAGGGCTCAGTTAA	2647
Db	2593	GTGTGATTTACACATCTCTCGCCTCATCTCTACGTCTCAAAACCCCGGTGTAGTTATAA	2652
QY	2648	AAATGAGTGAATTAATATCTCTAACGGCGCTAATTTTACCTTAAATTCCTCATGTTCCGCG	2707
Db	2653	ATTCTAAATACTTTTAATGTTTCAACAGGTCAGTTTAAGATTTAAAACTTTCAGGCTCAA	2712
QY	2708	GGCATGACGCTTTTAAATCAACAAAGACTTAACCATATAATGCAACCAATTCAAATTTCA	2767
Db	2713	CAAAAACTGCTTCTCAATAGAGAAAGATTAACTTTTAATGTCACC-----GGAGGCA	2766
QY	2768	GCCTCAGACAGACGAAGATGATTTTATGACGGGTACGCACGGAATGCCATCAATTCAA	2827
Db	2767	ACATAACATTTTGCRAAGTTGAAGCACCGCATGGAATGATTTGTTAAAGGCAATTTGAGCCA	2826
QY	2828	CCTACACATATCCATCTCTGGCGGTAATGTCACCCCTTGCTGGACAAAACTCAAGCAGCA	2887
Db	2827	AAAAAACATCAOCTTTGAAGGAGGTAACATCACCTTTGGCTCCAGGAAGCCGTACAG	2886
QY	2888	GCATTACGGGGAATATCTACTCGAGAAAGCACAATGTTAGCTTGAAGCCCAATAACG	2947
Db	2887	AAATCGAAGCAATGTTACTATCAATAACACGCTAACGTCAGTCTATCGGTCGGAAT	2946
QY	2948	COCTTAATCAGCAAAACATAAGGGATAGAGTTATAAACTTTGGCAGCTTGCTCGTTAATG	3007
Db	2947	TTGACAAACCATCAAAA-----ACCTTTAACTATTAAAGAAAGTGCATCATTAATAGCG	3000
QY	3008	GGAGTTTAAGTTTAACTGGCGAAATGCAGATATTAAAGCCAATCTCACTATTTCAGAAA	3067
Db	3001	GCACCTTACCCTGGAGGCAATATTGTCAATATACCCGGAATCTTACCGTTGAAGTA	3060
QY	3068	GGCCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATCACCGGCAATTTTACCAATA	3127
Db	3061	ACGCTAATTTCAAGCTATCACAAATTTTCACTTTTAATGTAGCGGCTTGTTTTGACACA	3120
QY	3128	ATGGCACTCCGGAATTAATATAACACAAGGAGTGTGTAATACTTGGCAATGTTACCAATG	3187
Db	3121	AAGGCAATTTCAATATTTCCATTGGCCAAAGAGGGCTCGCTTTAAAGACATTTGATAAT	3180
QY	3188	ATGTTGATTAAACATTTACCCTCAGCTTAACGCAACCAAGACGATCATCGGCGGAG	3247
Db	3181	CCAAGATTTAAGCATCACCACTCCAGCTCCACTTCCGCACTATTATAAGCGCA	3240
QY	3248	ATATAATCAAAAAAAGGAAGCTTAAATATTACAGACAGTAAATATGATGCTGAATCC	3307
Db	3241	ATATAACCAATAAAAACGGTGAATTTAAATATTACGAACGAAGTACTGACTGAAATGC	3300
QY	3308	AAATTTGGCGGCAATATCTCGAAAAAGGCAACCTCAGATTTCTTCCGATAAAATTA	3367
Db	3301	AAATTTGGCGGCGATGTCGCAAAAAGAGGTAATCTCAGGATTTCTTCTGACAAAAATCA	3360
QY	3368	ATATCACCACACAGATAACAATCAAAAAAGGTATTGATGGAGGAGCTCTAGTTTCAGATG	3427
Db	3361	ATATTACCAACACAGATAACAATCAAGGCGAGGTGTTGATGGGAGAAATTCGGATTACAGC	3420
QY	3428	CGACAAGTAATGCCCACTACTATTAAACCAAGAAATGGAATTCAGACAGAAGCACTAA	3487
Db	3421	CGACAAACATGCCAATCTCAACATTAACCAACCAAGAAATGGAATTAACGAAGACCTAA	3480
QY	3488	GTATTTCAGGTTTCAATAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTTAACATTG	3547
Db	3481	ATATTTCAGGTTTCAATAAAGCAGAGATTACAGCTAAAGATGGTAGTGAATTTAATCTATTG	3540
QY	3548	GCACAGTAATGACGGTTAACAGCGGTGCCGAAGCCAAAAACAGTAACCTTTTAACATGTTA	3607
Db	3541	GTAACACCAATAGTGCTGA---TGGTACTAATGCCAAAAAAGTAACCTTTTAACCAAGTTA	3597
QY	3608	AAGATTCAAAAATCTCTGCTCAGCGTCACATGTGCACACTAAATAGCAAGAGTGAACAAT	3667
Db	3598	AAGATTCAAAATCTCTGCTCAGCGTCACAAGGTGACACTACACAGAAAGTGGAAACAT	3657
QY	3668	CTAGCAGCAATGGCGGAGCGTGAAGCAATACGCAACGATACCGGCTTAATCTATTACTG	3727



```

Db 3658 CCGGTAGTAAATACAACTGAACATAGCAGTGACAAATAATGCGCGCTTAACATATCGATG 3717
QY 3728 CAAAAATGTAAGAGTAACAAAGATATATCTTCTCTCAAAACAGATAAATATCAGCGGT 3787
Db 3718 CAAAAATGTAAGAGTAACAAAGATATATCTTCTCTCAAAACAGATAAATATCAGCGGT 3777
QY 3788 C---GGAAAGAGTTACCAACAGAGCGCTCGACCATTAACGCAACAAATGCGCA- 3839
Db 3778 CAAAGTGGAGAAATATACCACTAAACAGAGTACAAACCATTAACGCAACCACTGGTAACGTGG 3837
QY 3840 ----- 3839
Db 3838 AGATAACCGCTCAAAACAGGTAGTATCTAGGTGGAATGAGTCCAGCTCTGGCTCTGTAA 3897
QY 3840 ----- 3839
Db 3898 CACTTACTGCAACCGAGGGCGCTCTTGCTGTGAAGCAATATTTTCGGGCAACACCGTTACTG 3957
QY 3840 ----- 3839
Db 3958 TTACTGCAATAGCGGTGCAATTAACACATTTGGCAGGCTCTACAATTAAGCAACCGAGA 4017
QY 3840 -AGCAAGTATTAACAAACAGGTGATATACAGCGGTACGATTTCCGGTACACAGGTAA 3988
Db 4018 GTGTAACCACTTCAAGTCAATACAGCGATATCGCGGTACGATTTCTGGTGGCACAGTAG 4077
QY 3899 GTGTTAGCGGCTGCTGATTTAAACACATTAATCCGGCTCAAAATTTGAAGCGCAATCGG 3958
Db 4078 AGGTTAAAGCAACCGAAAGTTTAAACACATCAATCCCAATTAAGCAACCAACAG 4137
QY 3959 GTGAGGCTTAATGTAACAAAGTGAACAGGTACAAATTTCCGGTACAAATTTCCGGTAAATACGG 4018
Db 4138 GCGAGGCTTAACAAAGTGAACAGGTACAAATTTCCGGTACAAATTTCCGGTAAATACGG 4197
QY 4019 TAAATGTTACGCAACCGCTGCGGATTTAAACAGTTGCGGATGCGGAGCAAAATTAATGCGA 4078
Db 4198 TAAATGTTACGCAACCGCTGCGGATTTAAACAGTTGCGGATGCGGAGCAAAATTAATGCGA 4257
QY 4079 CAGAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGAGTACTGAGCGGTTCTTA 4138
Db 4258 CAGAGGAGCTGCAACCTTAACCTATCATCATCGGCAAAATTAACCTACGAGCTAGTTTCA 4317
QY 4139 GCATCACTTCACTAAGGCTCAGTACAGCTCTTGCTCAGATGCTAGCATCGCAGGAA 4198
Db 4318 ACATTACTTCAAGGCTCAGTACAGCTCTTGCTCAGATGCTAGCATCGCAGGAA 4377
QY 4199 GCATTAATGCTGTAATGTGACATTAATATCTACAGGCACTTAAACAGCGTGGCAGGCT 4258
Db 4378 GTATTAATGCGGCAATGTGACACTAAATATCTACAGGCACTTAACTACCGTGAAGGTT 4437
QY 4259 CGGATATTAAGCAACAGCGGCACTTGTTTATTAACGCAAAAGATGCTTAAGCTAAATG 4318
Db 4438 CAAACATTAATGCAACAGCGGTAACCTTGTTTATTAACGCAAAAGATGCTTAAGCTAAATG 4497
QY 4319 GTGATGATCAGGTGATGATACAGAGTGAATGATGATGATGATGATGATGATGATGATG 4378
Db 4498 GCGCAGCATTTGGTAAACACACAGTGTGATGATGATGATGATGATGATGATGATGATG 4557
QY 4379 TGACTGCGGCAACCTCAAGCAGTGTGATATCTAGTGGGATTTAAACACAGTAAATGGGT 4438
Db 4558 TAATCGGCAACCTCAAGCAGTGTGATATCTAGTGGGATTTAAATCAACATTAATGGAT 4617
QY 4439 TAAATATCATTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 4498
Db 4618 TAAATATCATTTCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 4677
QY 4499 TGAATATATCCAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4558
Db 4678 TGAATATATCAACCGGTTAGCAGCGTGTGATGATGATGATGATGATGATGATGATGATG 4737
QY 4559 TTGAAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAACTTGGTCTAA 4618

```

```

Db 4738 TTGAGAAGGTAAAGATTTATCTGATGAAGAAAGAGCGTTAGCTAAACTTTGGAGTAA 4797
QY 4619 GTGCTGTACGTTTGTGTTGAGCCAAATATACAAATACAGTCAATATACAAATGAATTTA 4678
Db 4798 GTGCTGTACGTTTGTGTTGAGCCAAATATACAAATACAGTCAATATACAAATGAATTTG 4857
QY 4679 CAACGACCGCTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTCTCTCAAGTGTGA 4738
Db 4858 CAACGACCGCTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTCTCTCAAGTGTGA 4917
QY 4739 ATGCGGCACGAGTATGATACCAATTTCTGATGATGATGATGATGATGATGATGATG 4798
Db 4918 ATGCGGCACGAGTATGATACCAATTTCTGATGATGATGATGATGATGATGATGATG 4977
QY 4799 AGGTAGATTTTCATCTCTGCAATGAAGTCAATTTTATTTTCTGATTTTACTGTGTGGTTA 4858
Db 4978 AGGTAGATTTTCATCTCTGCAATGAAGTCAATTTTATTTTCTGATTTTACTGTGTGGTTA 5037
QY 4859 AAGTTTCAGTACGCGGCTTTTACCCATCTTGTAAAAAATACGAGAGTAAATATAAATGATTT 4918
Db 5038 AAGTTTCAGTACGCGGCTTTTACCCATCTTGTAAAAAATACGAGAGTAAATATAAATGATTT 5097
QY 4919 TTAACAGGTTATTTATG 4937
Db 5098 TTAACAGGTTATTTATG 5116

RESULT 11
AAAS2195
ID AAAS2195 standard; DNA; 5116 BP.
XX
AC AAAS2195;
XX
DT 11-SEP-2000 (first entry)
XX
DE Haemophilus influenzae strain 12 hmw1A gene, SEQ ID NO:66.
XX
KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; Nthi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; Otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis; ds.
XX
OS Haemophilus influenzae strain 12.
XX
FH Key Location/Qualifiers
FT CDS 351..4961
FT /tag= a
FT /product= "Haemophilus influenzae strain 12 hmw1A
XX protein"
XX
WO200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
PR 08-DEC-1998; 98US-0205942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
XX
P-PSDB; AAB01846.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
XX Example 16; Fig 28A-Q; 307pp; English.
XX

```

CC The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in *Escherichia coli*. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in *E. coli* (e.g., the T7 promoter) operably linked  
 CC to a modified hmWABC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene  
 CC clusters termed hmw1ABC and hmw2ABC. Each hmWABC operon comprises hmwA,  
 CC hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins  
 CC and the hmwB and hmwC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMWA  
 CC proteins. The modified hmWABC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwA genes (AA52175-A52198)  
 CC and HMWA proteins (AA01824-B01849) from the non-typeable H. influenzae  
 CC strains Jcy, K1, K21, LDC2, FWH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an hmwA gene from a non-typeable strain of H.  
 CC influenzae.

XX  
 SQ Sequence 5116 BP; 1766 A; 1008 C; 1037 G; 1305 T; 0 other;

Query Match 54.3%; Score 2680.2; DB 21; Length 5116;  
 Best Local Similarity 73.5%; Pred. No. 0;  
 Matches 3761; Conservative 0; Mismatches 1093; Indels 265; Gaps 16;

QY 41 ATGCAAAACACAAATACAAACCTTTTTCGAGTCTATATGCAAAATATTTTAAAAAAT 100  
 DB 41 ATGCAAAACACAAATACAAACCTTTTTCGAGTCTATATGCAAAATATTTTAAAAAAT 99  
 QY 101 AGTAAATTCGCCCATATAAATAGTAAATCTTTTCATCTTTTCATCTTTTCATCTTCATC 160  
 DB 100 AGTAAATTCGCCCATATAAATAGTAAATCTTTTCATCTTTTCATCTTTTCATCTTCATC 159  
 QY 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220  
 DB 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219  
 QY 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 280  
 DB 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 279  
 QY 281 GAATGAAGAGGAGCTGAACGAAACCAATGATAAGTAAATTTTAACTAACT 340  
 DB 280 GAATGAAGAGGAGCTGAACGAAACCAATGATAAGTAAATTTTAACTAACT 339  
 QY 341 TAGGAGAAATATGAACAAGATATATCTCTCAAAATTCAGCAACGCTCAATGCTTTGG 400  
 DB 340 TAGGAGAAATATGAACAAGATATATCTCTCAAAATTCAGCAACGCTCAATGCTTTGG 399  
 QY 401 TTGCTGTCTGAATTTGGCAGCGGGTTGTGACCAATTTCCACAGAAAAAGGCTTCCGCTAG 460  
 DB 400 TTGCTGTCTGAATTTGGCAGCGGGTTGTGACCAATTTCCACAGAAAAAGGCTTCCGCTAG 459  
 QY 461 TTATCATCTTTAGTGTACCACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTT 520  
 DB 460 TTGCTGTCTGAATTTGGCAGCGGGTTGTGACCAATTTCCACAGAAAAAGGCTTCCGCTAG 519  
 QY 521 TAGGTGAATCACTATTCCACAATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGAG 580  
 DB 520 TAGGTGAATCACTATTCCACAATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGAG 579  
 QY 581 TACAGGCAAGCCACTATGCAAGTAGATGGTAAATAAACCATTTATCCGCAACAGTGTG 640  
 DB 580 TACAGGCAAGCCACTATGCAAGTAGATGGTAAATAAACCATTTATCCGCAACAGTGTG 639

QY 641 ACGTATCATTAATTTGGAACAAATTTTAACTCGACCAAAATGAAATGGTGCAAGTTTATAC 700  
 DB 640 ACGTATCATTAATTTGGAACAAATTTTAACTCGACCAAAATGAAATGGTGCAAGTTTATAC 699  
 QY 701 AAGAAACAAACAACTCCGCGGTATTCACACCGTGTATTAACCAAAATCTCCCAATTA 760  
 DB 700 AAGAAACAAACAACTCCGCGGTATTCACACCGTGTATTAACCAAAATCTCCCAATTA 759  
 QY 761 AAGGATTTAGATTTTAAAGGACAAAGTCTTTTAAATCAACCAATGATATCAATAG 820  
 DB 760 AAGGATTTAGATTTTAAAGGACAAAGTCTTTTAAATCAACCAATGATATCAATAG 819  
 QY 821 GTAAGACGCAATTTTAAACACTTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 880  
 DB 820 GTAAGACGCAATTTTAAACACTTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 879  
 QY 881 AAAACATCAAGCGCGGTAAATTTTACCTTCGAGCAAAACAAAGATAAAGCGCTCGCTGAA 940  
 DB 880 AAAACATCAAGCGCGGTAAATTTTACCTTCGAGCAAAACAAAGATAAAGCGCTCGCTGAA 939  
 QY 941 TTGTGAATCAAGCGGTAAATTTTACTGTCTGTAAGACGCGAGTGTAAATCTTATTTGGTGCA 1000  
 DB 940 TTGTGAATCAAGCGGTAAATTTTACTGTCTGTAAGACGCGAGTGTAAATCTTATTTGGTGCA 999  
 QY 1001 AAGTGAACAAAGCGGTGTGATTTAGCGTAAATGTGCGAGCATTTCTTTACTCGCAGGCG 1060  
 DB 1000 AAGTGAACAAAGCGGTGTGATTTAGCGTAAATGTGCGAGCATTTCTTTACTCGCAGGCG 1059  
 QY 1061 AAAAATCACCATCAGCGATATAAATAACCAACCAATTTACTTACAGCAATTTGCGCGCGCTG 1120  
 DB 1060 AAAAATCACCATCAGCGATATAAATAACCAACCAATTTACTTACAGCAATTTGCGCGCGCTG 1119  
 QY 1121 AAAATGAAGCGGTCAATCTGGCGGATATTTTGCAGAAAGCGGTAAACATTAATGTCCGCTG 1180  
 DB 1120 AAAATGAAGCGGTCAATCTGGCGGATATTTTGCAGAAAGCGGTAAACATTAATGTCCGCTG 1179  
 QY 1181 CTGCCACTTTCGAAACCAAGTAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCG 1240  
 DB 1180 CTGCCACTTTCGAAACCAAGTAACTTCTGCTGATTTCTGTAAGCAAGATAAAGCG 1239  
 QY 1241 GCAATATTGTTCTTTCCGCCAAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTC 1300  
 DB 1240 GCAATATTGTTCTTTCCGCCAAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTC 1299  
 QY 1301 AAAATCAGCAAGCTTAAAGCGCGAGCTGATGATTACAGGCGGATAAAGTCAACATTA 1360  
 DB 1300 AAAATCAGCAAGCTTAAAGCGCGAGCTGATGATTACAGGCGGATAAAGTCAACATTA 1359  
 QY 1361 CAGGTGCAAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGAG 1420  
 DB 1360 CAGGTGCAAGTATTCGACCTTTTCAGGTAAAGAGGGGAGAACTTACCTTTGGCGGTGAG 1419  
 QY 1421 AGCGGCGGAGGTAAAAACCGCATTTCAATTTAGCAAGAAAAACCTCTTTAGAAAAAGCT 1480  
 DB 1420 AGCGGCGGAGGTAAAAACCGCATTTCAATTTAGCAAGAAAAACCTCTTTAGAAAAAGCT 1479  
 QY 1481 CAACCATCAATGTATCAGCAAGAAAAAGAGCGGCGCTATTTGTTGGGGGATATTG 1540  
 DB 1480 CAACCATCAATGTATCAGCAAGAAAAAGAGCGGCGCTATTTGTTGGGGGATATTG 1539  
 QY 1541 CGTTAATTGAGCGCAATATTAAACGCTCAAGGTAGTGGTGATATCGCTTAAACCGGTGCTT 1600  
 DB 1540 CGTTAATTGAGCGCAATATTAAACGCTCAAGGTAGTGGTGATATCGCTTAAACCGGTGCTT 1599  
 QY 1601 TTGTGGAGACATCGGGGCAATTTATTTTACCAATGACAGCAATGCAATTTGTTAAACAAAG 1660  
 DB 1600 TTGTGGAGACATCGGGGCAATTTATTTTACCAATGACAGCAATGCAATTTGTTAAACAAAG 1659  
 QY 1661 AGTGTGTCTAGACCTCTGATGATCAACAAATTTAAGCGCGAGACCCCTTTCCCAATATA 1720  
 DB 1660 AGTGTGTGTAGACCCGATGATGATCTATTATGACAGAAACAGCAGCAGCAATA 1719  
 QY 1721 CCGGTATAAATGATGAATTTCCCAACAGCGCGGTGAAGCAAGCAAGCCCTTAAAAAATA 1780

Db 1720 CTTGAGAGAGATGATACACGGGATCCGGGATAGTCCGACACCCCAAAACGAAACA 1779  
 Qy 1781 GCGNACTCAAAACAGCGTAAACAAATACAACTATTTCAAAATATCTGAAACAGCGCTGGA 1840  
 Db 1780 AAGAA---AAGACAAATTAACAAACAACTCTTGAGAGTATACATAAAAGAGTACCT 1836  
 Qy 1841 CAATGAATATACGGGATCAAGAAACTTACCGTTAATAGCTCAATCAACATCGGAAGA 1900  
 Db 1837 TTGTTAACTCACTGTAATCAACGATCTATGTCATAGCTCAATTAATTT---ATCCA 1893  
 Qy 1901 ACTCCCACTTAATCTCCATAGTAAGGTCAGCGTGGGAGCGGTTACAGATTGATGGAG 1960  
 Db 1894 ATGGCAGCTTAACCTTTGGAGTGAGGTCGAGCGGTGGCGGTTGAGATTAAACAAG 1953  
 Qy 1961 ATATTAC-----TTCTAAAGGCGGAAATTAACCAATTTATCTGGCGGATGG 2008  
 Db 1954 ATATTACCCGGTGATGATACAGAGGTGCAAACTTAACAATTTACTCAGCGCGTGG 2013  
 Qy 2009 TTGATGTTCAATAAATATACCTTTGATCAGGGTTTAAATATATACCGCGCTTCCG 2068  
 Db 2014 TTGATGTTCAATAAATATCTACTCGGGCGCAAGGTAACTAAACATTAACAGCTAAAC 2073  
 Qy 2069 TAGCTTTTGAAGGTGGAATTAACAAAGCAGCGGCGCAATGCTAAATGTCGCC 2128  
 Db 2074 AAGATATCGCTTTGAGAAGGAGCAACCAAGTC-----ATTACAGTTC 2118  
 Qy 2129 AGGCACTGTAACTTACAGGAGAGGAAAGATTTCAGGGGTACACAGCTATCTTTAA 2188  
 Db 2119 AAGGAGCTATTACC---TCAGGCAATCAAAAGGTTTAAATGATTAATGATCTCTCTAA 2175  
 Qy 2189 ACGGAACGGGTAAAGGTCCTGAATATCATTTTCATCAGTGAATAA-----TTTAAACCC 2239  
 Db 2176 ACGGCACTGCGACGGGACTGCAATTCACCACATAAAGAGCAATAAATAGCTATCACAA 2235  
 Qy 2240 ACATCTTAGTGACACAAATTAACATATCTGGGAATATAACAAT---TAACCAAACTACGA 2296  
 Db 2236 ATAAATTTGAAGGACCTTTAAATATTTTACGGGAAAGTGAACATCTCAATGGTTTACCTA 2295  
 Qy 2297 GAAAGAACCTCTGATTTGCAACACAGCCATGATTCGCACTGGAAGCTCAGTGCCTTA 2356  
 Db 2296 AAATGAAGTGGATATGATAAATTCAAAGGACGCACTTACTGGAATTTAACTCTCTTA 2355  
 Qy 2357 ATCT---AGAGACAGCGCAAAATTTTACCTTATTAATACATTTCAACCAATAGCAAG 2413  
 Db 2356 ATGTTCCGAGAGTGGCGAGTTAACTCTACTATGCTCCAGAGGAGCGATAGTGCAG 2415  
 Qy 2414 GCTTAAACACAGATATAGAGCTCTGAGGGTGAATTTTACGGGTTAAATGGCAACA 2473  
 Db 2416 GCACCTTACCCAGCTTATTAATTTAAACGGTATATCAATCAACAAAGACACT---ACCT 2472  
 Qy 2474 TGCTAATCAATCTCAAGAGGAGCGAAAGTTAATTTCAATTTAAACCAACAGAGAA 2533  
 Db 2473 TTAATGTTGAACGAAATGCAAGAGTCAACTTTGACATCAAGGCGCCCAATAGGGATAAATA 2532  
 Qy 2534 TGAACACAGCAACCTTTTACCAATTCGGTTTTAGCCAAATATCACAGCCACTGGTGGG 2593  
 Db 2533 AGTATCTAGTTTGAATTTAGCGATCAATTTAATGGAACAATTTTCAGTTTCGGGAGGGGGA 2592  
 Qy 2594 GCTCTGTTT-----TTTTGATATATATGCCAACCACTTCTGCGAGAGGCGTGAATTA 2647  
 Db 2593 GTGTTGATTTTACACTTCTCGCTCTAATGCTTAAAGTCCAAACCCCGGCTGATGTTATA 2652  
 Qy 2648 AAATGAGTGAATTAATCTCAACGCGCTAATTTTACCTTAAATTTCCCATGTCGCG 2707  
 Db 2653 ATTCTAAATCTTAAATGTTTCAACAGGTCAGTTTAAAGATTTAAACTTTCAGGCTCAA 2712  
 Qy 2708 GCGATGAGCTTTTAAATCAACAAAGACTTAACCAATAAATGCAACCAATTTCAATTTCA 2767  
 Db 2713 CAAAACCTGGCTTCAATAGAAAGATTTTAACTTTAAATGGCAACC-----GGAGGCA 2766  
 Qy 2768 GCCTCAGACAGCAAGAGATTTTATGACGGGTACGACGCAATGCGCATCAATTTCAA 2827

Db 2767 ACATAACACTTTTGAAGTTGAAGCCACGATGGNAATGATTTGGTAAAGCACTTTAGCCA 2826  
 Qy 2828 CCTACAACTATCCATTTCTGGCGGTAAATGTCACCCCTTGGTGGGACAAAACTCAAGCAGCA 2887  
 Db 2827 AAAAACAATACCTTTGAAGGAGTAACTACCTTTGGCTCCAGAGAAAGCCGTAAACAG 2886  
 Qy 2888 GCATTCGGGGAAATATCTATCTGAGAAAGCAGCAAAATGTTTACGCTAGAAGCAATAACG 2947  
 Db 2887 AAATCGAAGGCAATGTTACTATCAATAAAGCGTAAAGCTCACTCTTATCGGTTGCGGAT 2946  
 Qy 2948 CCCCTAATCAGCAAAACATAAGGGATAGAGTTATATAAACTTTGGCAGCTTGGCTTAAATG 3007  
 Db 2947 TTGACAACCATCAAAA-----ACCTTTAACTATTTAAAAAAGATGTCATCATTAATAGCG 3000  
 Qy 3008 GGAGTTTAAAGTTTAACTTGGCGAAATGAGATATTAAGGCAATCTCACTATTTTCAGAAA 3067  
 Db 3001 GCAACCTTACCCTGGAGCAATATTTGTCATATATAGCGGAAATCTTACCGTTGAAAGTA 3060  
 Qy 3068 GCGCCACTTTTAAAGAAAGACTAGAGATACCTTAAATATCACCGCAATTTTACCAATA 3127  
 Db 3061 ACGCTAATTTCAAGCTATCACAAATTTTCACTTTTAAATAGTGGCGCTTGTGTTGACAACA 3120  
 Qy 3128 ATGSCACTGCGCAATTAATATACACAGGAGTGGTAAACTTGGCAATGTTTACCAATG 3187  
 Db 3121 AAGCAATTTCAATATTTCCATGCAAGAGGAGGCTCGCTTTAAAGACATTTGATTAAT 3180  
 Qy 3188 ATGTTGATTTAAACATTTACCCTACGCTAAACGCAACCAAGAGCATATATGATGCTGAAATCC 3247  
 Db 3181 CCAAGAAATTTAAGCATCACCACTCCAGCTCCACTTTACCGCACTATTAATAAGCGCA 3240  
 Qy 3248 ATATAATCAACAAAAGGAAGCTTTAAATATTTACAGACAGTAAATATGATGCTGAAATCC 3307  
 Db 3241 ATATAACCAATTAACCGTGTATTTTAAATATTTACGAACGAGGTAGTATGATGAAATGC 3300  
 Qy 3308 AAATTTGGGCGCAATATCTCGCAAAAAGAGCAACCTCACGATTTCTCCGATAAAATTA 3367  
 Db 3301 AAATTTGGGCGCAATATCTCGCAAAAAGAGGTAATCTCACGATTTCTCTGACAAATCA 3360  
 Qy 3368 ATATCACCACACAGATTAACAAATCAAAAAGGTTATTTGATGGAGAGGACTCTAGTTCAGATG 3427  
 Db 3361 ATATTACCAACACAGATAACAATCAAGGCGGTGTTGATGGGAGAAATTCGGAATTCAGAG 3420  
 Qy 3428 CCACAGTATGCCAACCTTAATTAACCAAGAAATTAAGAAATGACAGAGACCTAA 3487  
 Db 3421 CGACAAACAACTGCTTAACTTAAACCAAGAAATTTGAATTAACGCAAGACCTAA 3480  
 Qy 3488 GTATTTTCAAGTTTCAATTAAGCAGAGATTACAGCAAGAGATGGTAGAGATTTTAACTATTG 3547  
 Db 3481 ATATTTTCAAGTTTCAATTAAGCAGAGATTACAGCTAAAGATGGTAGTATTAACTATTG 3540  
 Qy 3548 GCAACAGTAAATGACGGTAAACAGCGGTGGGAAAGCCAAACAGTAACCTTTTAAACAATGTTA 3607  
 Db 3541 GTAACACCAATAGTCTGTA---TGCTACTTAATGCCCCAAAAGTAACCTTTTAAACAGGTTA 3597  
 Qy 3608 AAGATTTCAAAAATCTCTGCTGACGGTCAATATGTCACACTTAAATAGCAAGTGAACAAAT 3667  
 Db 3598 AAGATTTCAAAAATCTCTGCTGACGGTCAATAGTTCACAGTACACAGCAAGTGAACAAAT 3657  
 Qy 3668 CTAGCAGCAATGGCGGAGCTGAAAGCAATAGCGCAACCAATACCGGCTTTAACTATTACTG 3727  
 Db 3658 CCGGTAGTAAATCAACACTGGAAGTACAGTGAACAATTAATGCGGCTTTAACTATCGATG 3717  
 Qy 3728 CAAAAATCTAGAAATTAACAAAGATATTTACTTCTCTCAAAACAGTAAATATCACCGCT 3787  
 Db 3718 CAAAAATCTAGAAATTAACAAAGATATTTACTTCTCACAAGACGATGAGCACTCTCTGCA 3777  
 Qy 3788 C---GGAAAAGTACCACACAGCGCTCGACCAATTAACGCAACAAATGCAAA----- 3839  
 Db 3778 AAGTGGAGAAATTTACCACTAAACAGGTACAAACCAATTAACCAACCACTTGGTAAAGTGG 3837  
 Qy 3840 ----- 3839  
 Db 3838 AGATAACCGCTCAACACAGGTAGTATCTTAGTGGAAATTTAGTTCAGTCCAGCTCTGCTCTGTAA 3897

QY 3840 ----- 3839  
Db 3898 CACTTACTGCAACCGAGGCGCTCTGCTGTAAGCAATATTTCGGGCAACACCGTTACTG 3957  
QY 3840 ----- 3839  
Db 3958 TTACTGCAATATACCGGTGCATTAACCACTTTGGCAGGCTCTACAAATTAAGGAACCGAGA 4017  
QY 3840 -AGCAAGTATTACAAACCAACAGGTGATATCAGCGGTACGATTTCCGGTAAACACGGTAA 3898  
Db 4018 GTGTAACCACTTCAAGTCAATCAGCGGATATCGCGGTACGATTTCTGGTGCACAGTAG 4077  
QY 3899 GTGTTACCGGCACTGGTATTTAACCACTAAATCCGGCTCAAAAATTGAAGGAAATCGG 3958  
Db 4078 AGTTAAAGCAACCGAAGTTTAAACCACTCAATTCAAATTTCAAAAATTAAGCAACACAG 4137  
QY 3959 GTGAGGCTAATGTAACAAAGTCAACAGGTACAATTTGGCGGTACAAATTTCCGGTAATACGG 4018  
Db 4138 GCGAGGCTAACGTAACAAAGTCAACAGGTACAATTTGGGTACGATTTCCGGTAATACGG 4197  
QY 4019 TAAATGTTAGCGGAACCGTGGCGATTTAACAGTTGGGAATGGCGCAGAAATTAATCGGA 4078  
Db 4198 TAAATGTTAGCGGAACCGTGGCGATTTAACAGTTGGGAATGGCGCAGAAATTAATCGGA 4257  
QY 4079 CAGAAGGAGTCAACCTTAAACGCAACAGGGAATACCTTGACTACTGAAGCCGGTCTA 4138  
Db 4258 CAGAAGGAGTCAACCTTAACTACATCATCGGCAATTAACCTACCGAAGCTAGTCCAC 4317  
QY 4139 GCATCACTTCAACTAAGGGTCAGTAGACCTCTTGGCTCAGAATGGTAGCATCGCAGAA 4198  
Db 4318 ACATTACTTCAACCAAGGGTCAGGTAATCTTTCAGCTCAGGATGCTAGCGTTGCAGAA 4377  
QY 4199 GCATTATGCTCTAATGTGACATTAATACACTACAGGCACCTTAACCCCGTGCAGGCT 4258  
Db 4378 GTATTATGCCCCAAATGTGACATTAATACACTACAGGCACCTTAACTACCGTGAAGGGTT 4437  
QY 4259 CGGATATTAACCAACAGCGGCACCTTGGTTTATTAACGCAAAAGATGCTAAGCTAAATG 4318  
Db 4438 CAACATTAATCAACCAAGCGGTACCTTGGTTTATTAACGCAAAAGATGCTAAGCTAAATG 4497  
QY 4319 GTGATGATCATGCTAGTAGACAGATGAATGCACTGTCAGTCAACGCAAGCGGCTGGTAGTG 4378  
Db 4498 GCGCAGCATTTGGTAAACACAGATGTTAAATCAACCAAGCAATGGCTCCGGCAGCG 4557  
QY 4379 TGACTGCGCAACCTCAAGCAGGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGT 4438  
Db 4558 TAATCGCGCAACCTCAAGCAGAGTGAACATCACTCGGGATTTAATCACAATAAATCGAT 4617  
QY 4439 TAAATATCAATTCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGCAAGCAATTCAGG 4498  
Db 4618 TAAATATCAATTCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGCAAGCAATTCAGG 4677  
QY 4499 TGAATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCAAGCGGCTCC 4558  
Db 4678 TGAATATATCAATTCGAAAGATGTTAGCAAGTGTAGAAAGTAAATTTGAAGCAAGCGGCTCC 4737  
QY 4559 TCGAAAGTAAAGATTTATCTGATGAAGAAACAGAAACATTAAGTAACTGGGTGTA 4618  
Db 4738 TCGAAAGTAAAGATTTATCTGATGAAGAAAGAGAGCGGTAACTAACTGGAGTAA 4797  
QY 4619 GTCTGTACGTTTTGTTAGCCAAATTAACAATTTACAGTCAATACACAAATGAATTTA 4678  
Db 4798 GTCTGTACGTTTTTATGAGCCAAATTAACAATTTACAGTCAATACACAAATGAATTTG 4857  
QY 4679 CAACGAGCCGTCAGTCAAGTGAATTTCTGAAGTAGAGCGGTCTTCTCAAGTGGTA 4738  
Db 4858 CAACGAGCCGTCAGTCAAGTGAATTTCTGAAGTAGAGCGGTCTTCTCAAGTGGTA 4917  
QY 4739 ATGGCGCAGATGATGTACCAATTTGCTGACGATGACAGCGGTAGTCAAGTAAATGACA 4798  
Db 4918 ATGGCGCAGCGTGTGGTGTATATCGCTGTATACGGCGGTAGCGGTCAAGTAAATGACA 4977

QY 4799 AGGTAGATTTTCATCTCGCAATGAAGTCATTTTATTTTCGTATTATTTACTGTGCGTTA 4858  
Db 4978 AGGTAGATTTTCATCTCGCAATGAAGTCATTTTATTTTCGTATTATTTACTGTGCGTTA 5037  
QY 4859 AAGTTCAGTACGGGCTTTACCCATCTTGTAATAAATACGAGAAATACAAATGAATTT 4918  
Db 5038 AAGTTCAGTACGGGCTTTACCCATCTTGTAATAAATACGAGAAATACAAATGAATTT 5097  
QY 4919 TTAACAGGTATTATTATG 4937  
Db 5098 TTAACAGGTATTATTATG 5116  
RESULT 12  
AAQ49508  
ID AAQ49508 standard; DNA; 9220 BP.  
XX AC AAQ49508;  
XX DT 26-APR-1994 (first entry)  
XX DE Gene cluster for high molecular weight protein 1 (HWW1).  
XX KW HWW: high molecular weight protein; virus; vaccine; influenza;  
XX KW epitope; immunity; haemophilus influenzae; gene cluster; ss.  
XX OS Haemophilus Influenzae.  
XX FH Key  
XX FT CDS Location/Qualifiers  
FT CDS 351..4958  
FT /tag= a  
FT /product= High molecular weight protein 1.  
FT 5114..6748  
FT /tag= b  
FT /note= "One of a gene cluster for high molecular  
FT weight protein 1"  
FT CDS 7062..9011  
FT /tag= c  
FT /note= "One of a gene cluster for high molecular  
FT weight protein 1"  
FT repeat\_unit 2351..2380  
FT /tag= d  
FT /label= "Possible error in sequence. Alternative  
FT sequence for this region is CTTAAATGTT  
FT TCCGAGAGTG CCGAGTTTAA (See AAR41725)"  
XX W09319090-A.  
XX PD 30-SEP-1993.  
XX PF 16-MAR-1993; 93WO-US02166.  
XX PR 16-MAR-1992; 92GB-0005704.  
XX PA (BARE/) BARENKAMP S J.  
XX PA (INRM ) INSRM INST NAT SANTE & RECH MEDICALE.  
XX PI Barenkamp SJ;  
XX DR WPI: 1993-320683/40.  
XX DR P-PSDB; AAR41725, AAR41726, AAR41727.  
XX PT High molecular weight surface proteins - of non-typeable  
XX PT haemophilus which exhibit immunogenic properties  
XX PS Claim 9; Figure 6; 100pp; English.  
XX CC The isolation and purification of the high molecular weight protein  
XX CC enables the identification of the major protective epitopes of the  
XX CC protein by conventional epitope mapping. These epitopes can then be  
XX CC synthesised using standard techniques and incorporated into fully  
XX CC synthetic or recombinant vaccines.

XX	Sequence	9220 BP; 3101 A; 1861 C; 1805 G; 2453 T; 0 other;
SQ	Query Match	54.2%; Score 2677; DB 14; Length 9220;
	Best Local Similarity	73.4%; Pred. No. 0;
	Matches 3759; Conservative	0; Mismatches 1095; Indels 265; Gaps 16;
QY	41	ATGCAACAACAATACAAACACCTTTTTCGAGCTATATGCAAAATATTTTAAAAAAT 100
DB	41	ATGCAACAACAATATACAAACCTTTTTCGAGCTATATGCAAAATATTTT-AAAAAT 99
QY	101	AGTATAATCCGCCATATAAATGATATAATCTTTTCATCTTTTCATCTTTTAACTTTTCATC 160
DB	100	AGTATAATCCGCCATATAAATGATATAATCTTTTCATCTTTTCATCTTTTAACTTTTCATC 159
QY	161	TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220
DB	160	TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219
QY	221	ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 280
DB	220	ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 279
QY	281	GAATGAAGAGGAGCTGAACGAACCAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGG 340
DB	280	GAATGAAGAGGAGCTGAACGAACCAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGG 339
QY	341	TAGGAGAAATATGAACAGATATATCGCTCAAAATTCAGCAAAAGCGCTGAATGCTTTGG 400
DB	340	TAGGAGAAATATGAACAGATATATCGCTCAAAATTCAGCAAAAGCGCTGAATGCTTTGG 399
QY	401	TTGCTGTCTGCAATTTGCGAGGGTTGTGACCAATTCACAGAAAGAGGCTTCGCGCTATG 460
DB	400	TTGCTGTCTGCAATTTGCGAGGGTTGTGACCAATTCACAGAAAGAGGCTTCGCGCTATG 459
QY	461	TTTACTATCTTTAGGCTTAACCACTTTAGCGTTTAAAGCCACTTTTCGCTATGTTACTACT 520
DB	460	CTGCTCGCATGAAAGTGGCTCACTTAGCGTTTAAAGCCACTTTTCGCTATGTTACTACT 519
QY	521	TAGGTGTAACTATCTTACCAATCTGTTTGTAGCAAGCGCTTACAGGAATGGATGTAG 580
DB	520	TAGGTGTAACTATCTTACCAATCTGTTTGTAGCAAGCGCTTACAGGAATGGATGTAG 579
QY	581	TACAGCGCACAGCCACTATGCAAGTAGATGGTAATAAAACCAATTCACGCAACAGTGTG 640
DB	580	TACAGCGCACAGCCACTATGCAAGTAGATGGTAATAAAACCAATTCACGCAACAGTGTG 639
QY	641	ACGCTATCATTAATTTGAAACAATTTACATCGACCAAAATGAATGGTGGCTTTTAC 700
DB	640	ACGCTATCATTAATTTGAAACAATTTACATCGACCAAAATGAATGGTGGCTTTTAC 699
QY	701	TAGAAACAACAATCCGCGCTATTCACCGCTTTACATCTTAACCAATTCGCAATTTAA 760
DB	700	TAGAAACAACAATCCGCGCTATTCACCGCTTTACATCTTAACCAATTCGCAATTTAA 759
QY	761	AGGGATTTTATCTTAACGACAAAGCTTTTAAATCAACCCAAATGGTATCAACAATAG 820
DB	760	AGGGATTTTATCTTAACGACAAAGCTTTTAAATCAACCCAAATGGTATCAACAATAG 819
QY	821	GTAAAGACGCAATTAATTAACACTAATGGCTTTACGGCTTTACGCTAGACATTTCTAACG 880
DB	820	GTAAAGACGCAATTAATTAACACTAATGGCTTTACGGCTTTACGCTAGACATTTCTAACG 879
QY	881	AAAACATCAAGCGCGTAAATTTACCTTCGAGCAAAACCAAGATAAAGCGCTCCCTGAAA 940
DB	880	AAAACATCAAGCGCGTAAATTTACCTTCGAGCAAAACCAAGATAAAGCGCTCCCTGAAA 939
QY	941	TTGTGAATCCAGGTTTAACTCTGCTGAAGAGCGCAAGTGAATCTTATTTGGTGCCA 1000
DB	940	TTGTGAATCCAGGTTTAACTCTGCTGAAGAGCGCAAGTGAATCTTATTTGGTGCCA 999
QY	1001	AAGTGAACCAAGGCTGTGATTAAGCTTAATGCTGACGATTTCTTTTACTCGCAGGGC 1060

DB	1000	AAGTGAACCAAGGCTGTGATTAAGCTTAATGCTGACGACATTTCTTTACTCGCAGGGC 1059
QY	1061	AAAAATCACCANTCAGCGATATATTAACCAACCACTTACTTACAGATTTGCGCGCGCTG 1120
DB	1060	AAAAATCACCANTCAGCGATATATTAACCAACCACTTACTTACAGATTTGCGCGCGCTG 1119
QY	1121	AAATGAAGCGGCTCAATCTGGCGATATTTTCCAAAGCGGTAACATTAATTTGTCGCTG 1180
DB	1120	AAATGAAGCGGCTCAATCTGGCGATATTTTCCAAAGCGGTAACATTAATTTGTCGCTG 1179
QY	1181	CTGCCACTATTTCCAAACCAAGGTAACCTTTCTGCTGATTTCTGTAAGCAAGATAAAAAGC 1240
DB	1180	CTGCCACTATTTCCAAACCAAGGTAACCTTTCTGCTGATTTCTGTAAGCAAGATAAAAAGC 1239
QY	1241	GCAATATTTCTTTCCGCCAAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1300
DB	1240	GCAATATTTCTTTCCGCCAAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTC 1299
QY	1301	AAATCAGCAAGCTAAAGCGCGCAAGCTGATGATTACAGCGGATTAAGTCACATTAATAA 1360
DB	1300	AAATCAGCAAGCTAAAGCGCGCAAGCTGATGATTACAGCGGATTAAGTCACATTAATAA 1359
QY	1361	CAGTGCAGTTATCGACCTTTCAAGTAAAGAGGGGAGAACTTACCTTGGCGGTGACG 1420
DB	1360	CAGTGCAGTTATCGACCTTTCAAGTAAAGAGGGGAGAACTTACCTTGGCGGTGACG 1419
QY	1421	AGCGCGGCAAGGTAAAGCGGATTTCAATAGCAAGAAACCTCTTTAGAAAAAGGCT 1480
DB	1420	AGCGCGGCAAGGTAAAGCGGATTTCAATAGCAAGAAACCTCTTTAGAAAAAGGCT 1479
QY	1481	CAACCATCAATGTATCAGGCAAAAGAGCGGAGCGCTATTGTGTGGCGGATATTG 1540
DB	1480	CAACCATCAATGTATCAGGCAAAAGAGCGGAGCGCTATTGTGTGGCGGATATTG 1539
QY	1541	CGTTAATTTAGCGCAATATTAACGCTCAAGGTAGTGTGATATCGTAAAAACCGGTGTT 1600
DB	1540	CGTTAATTTAGCGCAATATTAACGCTCAAGGTAGTGTGATATCGTAAAAACCGGTGTT 1599
QY	1601	TTGTGAGACATCGGGGCTATTTATTCATTCAGCAAGCAATCAATTTGTTAAAAACAAAG 1660
DB	1600	TTGTGAGACATCGGGGCTATTTATTCATTCAGCAAGCAATCAATTTGTTAGCCTCAAG 1659
QY	1661	AGTGTGTGTAGACCTGTATGATTAACATTAAGCGGAGACCCCTTCGCAATTAATA 1720
DB	1660	AGTGTGTGTAGACCTGTATGATTAACATTAAGCGGAGACCCCTTCGCAATTAATA 1719
QY	1721	CCGGTATAATGATGATTTCCCAACAGGCAACCGGTGAGCAAGCGGACCTTAATAAATA 1780
DB	1720	CTTCAGAAAGAGATGATATACCGGATTCGGGAATAGTCCGACGACCCCAACGAAACA 1779
QY	1781	GCBAATCAAAACAGCTTAACCAATACAACTATTTCAATATCTGAAACACGCTGGA 1840
DB	1780	AAGAA--AAGAAACATTAACAAACAACTCTTGAGAGTATACTTAAACAAAGGATACCT 1836
QY	1841	CAATGAATATAACCGCATCAAGAAAACTTACCGTTAATAGCTCAATCAACATCGGAAGA 1900
DB	1837	TTGTTAACATCACTGTATCAACGATCTATGTCAATAGCTCAATTAATTT--ATCCA 1893
QY	1901	ACTCCCACTTAATTTCTCCATAGTAAAGGTCAGGTCGCGGAGGCTTCAGATGATGGAG 1960
DB	1894	ATGCAAGCTTAACTCTTTGAGTGAAGGTCGAGGCGGTGCGCGCTTGAGATTAACAAG 1953
QY	1961	ATATTAAC-----TTCTTAAAGCGGAAATTTAAACCATTTTATTTCTGCGCGATGGG 2008
DB	1954	ATATTAACCGGCTGATGATACCAAGGTTGCAAACTTACAAATTTTACTCAGCGCGCTGGG 2013
QY	2009	TTGATGTTCAAAAAATTTACGCTTTGATCAGGGTTTTTTTAAATATATACCGCGCTTCCG 2068
DB	2014	TTGATGTTCAAAAAATTTCTCCTCGGGCGCAAGGTAACATAACATTAACAGCTAAAC 2073
QY	2069	TAGCTTTTGAAGGTGGAATAACAAAGCACGCGACCGCAAAATGCTAAAAATTTGTCGCC 2128
DB	2074	AAGATATCGCTTTGAGAAAGGAAACCAAGTC-----ATTACAGGTC 2118



QY 2129 AGGGCACTGTAACCAATTACAGGAGAGGGAAGATTTTACGGCTAACACAGTATCTTTAA 2188  
DB 2119 AAGGCACTATTACC---TCAGCAATCAAAAGGTTTAGATTTAATAATGTCCTCTAA 2175  
QY 2189 ACGGAACGGGTAAAGTCTCTGAATATCATTTTCATCAGTGAATAA-----TTTAAACC 2239  
DB 2176 ACGGCACTGGCAGCGGACTGCAATTCACCACTAAAGAACCATAAATACGGTATCACAA 2235  
QY 2240 ACAATCTTAGTGGCAATTAACATATCTGGGAATATAACAAT---TAACCAAACTACGA 2296  
DB 2236 ATAAATTTGAAGGACTTTAAATATTTTACGGGAAAGTGAACATCTCAATGGTTTACCTA 2295  
QY 2297 GAAAGAACACCTCTGATTGGCAACACGACCATGATTTCGACATGGAACGTCAGTGTCTTTA 2356  
DB 2296 AAAATGAAAGTGGATATGATAAATTCAAAGGACGCATTTACGGAAATTAACCTCGAAG 2355  
QY 2357 ATCTAGAGACAGCGCAATTTTACCTT---TATTAAATACATTTCAAGCAATAGCAAG 2413  
DB 2356 TGGATATGATAAATTCAAAGGACGCCCTCACTATTGACTCCAGAGGAAGCGATAGTCAG 2415  
QY 2414 GCTTAAACACACAGATATAGAAGCTCTGAGGGTGAATTTTAAACGGCGTAATATGCAACA 2473  
DB 2416 GCACACTTACCCAGCTTATAATTTAAAGGTATATCATTCACAAAGACACT---ACCT 2472  
QY 2474 TGTCAATCTCAAGAGAGGGAAGTTAAATTTCAAAATTAACCAACCAAGCAACA 2533  
DB 2473 TTAATGTTGAAGCAATGCAAGAGTCAACTTTGACATCAAGGACCAATAGGGATATAA 2532  
QY 2534 TGAACACAGCAACCTTTTACCAATTCGGTTTATAGCAATATACAGCCACTGCTGGG 2593  
DB 2533 AGTATCTAGTTGAATTAACGATCATTTAATGGAAACATTTTCAGTTTCGGGAGGGGA 2592  
QY 2594 GCTCTGTTT-----TTTGTGATATATGCAACCACTCTGCGAGAGGGCTGAGTTAA 2647  
DB 2593 GTGTTGATTTTCACTCTCGCCCTCATCTTAAGCTGCAACCCCGCTGAGTTATAA 2652  
QY 2648 AAATGAGTGAATTAATCTCTAACGGCGTAAATTTTACCTTAAATTTCCCATGTTTCGG 2707  
DB 2653 ATTCTAAATACTTTAATCTTCAACAGGTCAGTTTAAAGATTTAAACTTCAGGCTCAA 2712  
QY 2708 GCGATGAGCTTTTAAATCAACAAGACTTAAACCATAAATGCAACCAATTTCAATTTCA 2767  
DB 2713 CAANAATCGGCTCTCAATAGAGAAGATTTAACTTTTAAATGCCACC-----GGAGCA 2766  
QY 2768 GCCTCAGACAGACGAAGATGATTTTATGAGGGTACGACGCAATGCCATCAATTTCAA 2827  
DB 2767 ACATAAACACTTTTGAAGTGAAGGACCGATGGAATGATTGGTAAAGGCATTTGAGCCA 2826  
QY 2828 CCTACACATATCCATTTCTGGCGGTATGTCACCCCTTGGTGGACAAACTCAAGCACA 2887  
DB 2827 AAAAACAATAACCTTTTGAAGGAGGTAAAGTGAAGTTTGGCTCCAGGAAGCCGTTAAG 2886  
QY 2888 GCATTACGGGAATATTACTTCGAGAAGCAGCAAAATGTTACGCTAGAAGCAATAACG 2947  
DB 2887 AAATCGAAGGCAATGTTACTATCAATAACAACGCTAACGTCACCTTATCGTTCGGATT 2946  
QY 2948 CCCCTAATCAGCAAAACATAGAGATAGATTAATAAATTTGCGAGCTTGGCTGTTAATG 3007  
DB 2947 TTGACACCATCAAAA-----ACCTTTAACTATTAATAAAGATGTCTATCAATTAATAGCG 3000  
QY 3008 GGAGTTTAAAGTTTAACTGGCGAAATGAGATATTAAGAGCAATCTCACTATTTTACAGAA 3067  
DB 3001 GCAACCTTACCGCTGGAGGCAATATTGTCATATAGCGGGAATCTTACCGTTGAAAGTA 3060  
QY 3068 CGCCACTTTTAAAGGAAGACTAGATACCCCTAAATATACCGGCAATTTTACCAATA 3127  
DB 3061 ACGTAAATTTTCAAGCTATCAAAATTTTCACTTTTAAATGAGGGCGCTGTTTACCAACA 3120  
QY 3128 ATGCACTGCCGAATTAATAACACAGGAGGTGGTAAATTTTGGCAATGTTTACCAATG 3187  
DB 3121 AAGGCAATCAAAATTTTCCATTTGCCAAGGAGGGCTCGCTTTTAAAGACATTTGATAAT 3180

QY 3188 ATGGTGAITTTAAACATTTACCCTACGCTAAACGCAACCAAGACATCATCGGGGAG 3247  
DB 3181 CCAAGAAATTTAAGCATCACCACCACTCCAGCTCCACTTACCCTACTATTATAAGCGCA 3240  
QY 3248 ATATATATCAACAAAGAAAGCTTAAATATTACAGACAGTATAATATGATGCTGAAATCC 3307  
DB 3241 ATATATACCAATTAACAAAGGCTGATTTAAATATTACGAAACGAAGGTAGTACTGAAATGC 3300  
QY 3308 AAATTTGGCGGCAATATCTCGCAAAAGAAAGCAACCTCACGATTTCTCCGATATAAATTA 3367  
DB 3301 AAATTTGGCGGCTGCTCGCAAAAGAAAGTAACTCACGATTTCTTCTGACAAATCA 3360  
QY 3368 ATATCACCACACAGATAACAATCAAAAGGTTATGATGAGAGGACTCTAGTTTCAGATG 3427  
DB 3361 ATATTACCACACAGATAACAATCAAGGAGGTGTTGATGGGAGAATTCGGATTCAGAGC 3420  
QY 3428 CGCAAGTAAATCCCACTTAACTATTAAACCAAGAAATTTGAAATTTGACAGAGACCTAA 3487  
DB 3421 CGCAAAACAATGCCAATCTAACCATTTAAACCAAGAAATTTGAAATTTAACGAGACCTAA 3480  
QY 3488 GTATTTCAAGTTTCAATTAAGCAGAGATTACAGCCAAAGATGGTAGAGATTTTAACATTG 3547  
DB 3481 ATATTTCAGTTTCAATTAAGCAGAGATTACAGCTAAAGATGGTAGATTTTAACATTG 3540  
QY 3548 GCACAGTAATGACGGTAACAGCGGTGCGGAAAGCCAAACAGTAACCTTTTAAACAATGTTA 3607  
DB 3541 GTAACACCAATAGTGTGA---TGGTACTAATGCCAAAGTAACCTTTTAAACAGGTTA 3597  
QY 3608 AAGATTTCAAAATCTCTGCTGACGGTCAACATGTGACACTAAATAGCAAGTGAACAT 3667  
DB 3598 AAGATTTCAAAATCTCTGCTGACGGTCAAGGTGACACTACACAGCAAGTGAACAT 3657  
QY 3668 CTAGCAGCAATGGCGAGCGTGAAGCAATAGCGCAACGATACCGGCTTTAAGCTTACTG 3727  
DB 3658 CCGGTAGTAAATAACACACATAGAGTGAAGTGAAGTAAATGCGGCTTTAAGCTTACTG 3717  
QY 3728 CAAAAATGTAGAAGTAAACAAAGATATTACTTCTCAAAACAGTAAATATACCGGCT 3787  
DB 3718 CAAAAATGTACAGTAAACAAATATTACTTCTCAAAACAGTGAAGCTTCTGCGCA 3777  
QY 3798 C---GGAAAGGTTACCAACAGAGGCTCGACCATTAACGCAACAAATGGCAA----- 3839  
DB 3778 CAAAGTGAAGAAATTAACCACTAAACAGGTAACCACTTAACGCAACCACTGTAAGCTG 3837  
QY 3840 ----- 3839  
DB 3838 AGATAACCCCTCAACACAGTAGTATCTTAGGTGGAATTCAGTCCAGCTCTGCTCTGTA 3897  
QY 3840 ----- 3839  
DB 3898 CACTTACTGCAACCGAGGCGCTCTTGTGTAAAGCAATTTTTCGGGCAACACCGTTACTG 3957  
QY 3840 ----- 3839  
DB 3958 TTACTGCAATAGCGGTGCATTAACCACTTTGGCAGGCTCTACAATTTAAAGGAACCGAGA 4017  
QY 3840 -AGCAAGTATTACAAACCAACAGAGTGATATCAGCGGTACGATTTCCGGTAACACGGTAA 3898  
DB 4018 GTGTAAACCACTTCAAGTCAATCAGCGGATAGCGGCTGACGATTTCTGCTGCGCAGTAG 4077  
QY 3899 GTGTAGCGGCACTGCTGATTTTAAACCACTAAATTCGGCTCAAAATTTGAAGGCAATCGG 3958  
DB 4078 AGGTTAAGCAACCGAAGTTTAAACCACTCAATCAATTTCAAAATTTAAAGCAACACAG 4137  
QY 3959 GTGAGGCTAATGTAACAAGTGCACAGGTACATTTGGGGGTACAATTTCCGGTAATACGG 4018  
DB 4138 GCGAGGCTTAAGCTAAACAAGTGCACAGGTACAATTTGGGTACGATTTCCGGTAATACGG 4197  
QY 4019 TAAATGTTACGCAACCGCTGCGGATTTTAAACAGTGGGAATGGCGCAGCAAAATTTAATGCA 4078  
DB 4198 TAAATGTTACGCAACCGCTGCGGATTTTAAACAGTGGGAATGGCGCAGCAAAATTTAATGCA 4257  
QY 4079 CAGAGGAGCTGCAACCTTAAACCGCAACAGGGAATACCTTGTACTGACTACTGAACCGGTTCTA 4138

```

Db 4258 CAGAGGAGCTGCAACCTTAACTACATCATCGGCAAAATAACTACCGAAGTAGTTAC 4317
Qy 4139 GCATCAGTCAACTAAGGTCAGGTAGACCTCTTGGCTCAGAAATGGTAGCTCGCAGGAA 4198
Db 4318 ACATTACTTCAGCAGGCTCAGGTAAATCTTTAGCTCAGGATGGTAGCTCGCAGGAA 4377
Qy 4199 GCATTAATGCTGCTAATGTCAGTAATAATACATACAGGCACTTAACACCGTGGCAGCT 4258
Db 4378 GTATTAAATGCCGCCAATGACACATAATACTACAGGCACTTAATACTACCGTGAAGGTT 4437
Qy 4259 CGGATATTAAGCAACACCGCGCACCTTGGTTATTAAACGCAAAAGATGCTAAGCTAAATG 4318
Db 4438 CAACATTAATGCAACACCGGTACCTTGGTTATTAAACGCAAAAGACGCTAAGTAAATG 4497
Qy 4319 GTGATGCAATGAGGTGATAGTACAGAAGTGAATGCAATCAACGCAAGCGGCTCTGGTAGTG 4378
Db 4498 CGCGACGATTTGGTAAACACACAGTGGTAAATGCAACCAACGCAAAATGGCTCCGCGAGCG 4557
Qy 4379 TGACTGCGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGT 4438
Db 4558 TAATCGCACACCTCAAGCAGTGAATCACTGGGATTTAAACACAGTAAATGGT 4617
Qy 4439 TAAATATCATTTGCAAGATGTTAGCAAGTGTAGAGAGTAAATTTGAAGCGAAGGAAATGAGG 4498
Db 4618 TAAATATCATTTCAAAAACGATTAACACCGTACTGTTAAAGCGGTAAATTTGATG 4677
Qy 4499 TGAATATATTCAGCAGGCTGTAGCAAGTGTAGAGAGTAAATTTGAAGCGAAGGAAATGAGG 4558
Db 4678 TGAATATATTCAGCAGGCTGTAGCAAGTGTAGAGAGTAAATTTGAAGCGAAGGAAATGAGG 4737
Qy 4559 TTGAAAAAGTAAAGATTTATCTGTAGTGAAGAGAGAGAAACATTTAGCTAAACCTTGGTGTA 4618
Db 4738 TTGAGAGGTTAAAGATTTATCTGTAGTGAAGAGAGAGAAACATTTAGCTAAACCTTGGTGTA 4797
Qy 4619 GTGCTGTACGTTTGTGAGCAAAATAATACAAATTTACAGTCAATACACAAAATGAATTTA 4678
Db 4798 GTGCTGTACGTTTGTGAGCAAAATAATACAAATTTACAGTCAATACACAAAATGAATTTG 4857
Qy 4679 CAACAGACCGTCAAGTGAATTTCTGGAAGTGAAGCGGCTTTCTCAAGTGTA 4738
Db 4858 CAACAGACCGTCAAGTGAATTTCTGGAAGTGAAGCGGCTTTCTCAAGTGTA 4917
Qy 4739 ATGGCGCACGAGTGTACCAATGTGTGAGTGAAGTGAAGCGGCTTTCTCAAGTGTA 4798
Db 4918 ATGGCGCACGAGTGTACCAATGTGTGAGTGAAGTGAAGCGGCTTTCTCAAGTGTA 4977
Qy 4799 AGGTAGATTTTCATCTGCAATGAAGTCAATTTATTTCTGATTTTACTGTGGGTTA 4858
Db 4978 AGGTAGATTTTCATCTGCAATGAAGTCAATTTATTTCTGATTTTACTGTGGGTTA 5037
Qy 4859 AAGTTCAGTACGGCTTTTACCCATCTTGTAAATAATTTACGAGAAATACAAATAAGATATT 4918
Db 5038 AAGTTCAGTACGGCTTTTACCCATCTTGTAAATAATTTACGAGAAATACAAATAAGATATT 5097
Qy 4919 TTAACAGTATTATATG 4937
Db 5098 TTAACAGTATTATATG 5116

```

RESULT 13

AAT90996

ID AAT90996 standard; DNA; 9171 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

OS
XX
FH Key Location/Qualifiers
FT CDS 351..4958
FT /*tag= a
FT /*note= "hmw1 gene"
FT CDS 5114..6748
FT /*tag= b
FT /*note= "ORF-b"
FT CDS 7062..9011
FT /*tag= c
FT /*note= "ORF-c"
PN W09736914-A1.
XX
XX 09-OCT-1997.
XX 01-APR-1997; 97WO-US04707.
XX 01-APR-1996; 96US-0617697.
XX (BARE/) BARENKAMP S J.
XX Barenkamp S J;
XX WPI; 1997-503038/46.
XX P-PSDB; AAW30293.
XX High molecular weight proteins of non-typeable Haemophilus
XX influenzae - useful for vaccine production
XX Disclosure; Page 78-82; 183pp; English.
XX This nucleic acid comprises a gene cluster for the hmw1 gene (see
XX also AAT90994) that encodes high molecular weight surface protein
XX HMW1 (see AAW30293) of non-typeable Haemophilus influenzae strain 12.
XX In addition to the hmw1 gene there are 2 additional downstream
XX open reading frames that are required for the correct processing
XX and secretion of the hmw1 gene product. The ORF-b derived amino
XX acid sequence demonstrates similarity with the derived amino acid
XX sequences of 2 genes which encode proteins required for secretion
XX and activation of haemolysins of P. mirabilis and S. marcescens.
XX HMW proteins (see AAW30291-94) can be used in vaccines, as immunogens
XX for preparation of antibodies and as antigens for detection of
XX these antibodies.
SQ Sequence 9171 BP; 3085 A; 1855 C; 1794 G; 2437 T; 0 other;
Query Match 52.0%; Score 2565.4; DB 18; Length 9171;
Best Local Similarity 72.4%; Pred. No. 0;
Matches 3708; Conservative 0; Mismatches 1096; Indels 315; Gaps 17;
Qy 41 ATGACAAACAACAATTAACAACACCTTTTTCGAGTCTATATGCAATAATTTTAAAAAAT 100
Db 41 ATGACAAACAACAATTAACAACACCTTTTTCGAGTCTATATGCAATAATTTT-AAAAAT 99
Qy 101 AGTATAAATCCGCCATATAAATAATGTTAATCTTTTCATCTTTTCATCTTTTCATCTTC 160
Db 100 AGTATAAATCCGCCATATAAATAATGTTAATCTTTTCATCTTTTCATCTTTTCATCTTC 159
Qy 161 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 220
Db 160 TTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 219
Qy 221 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTC 280
Db 220 ATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTC 279
Qy 281 GAATGAAGAGGAGCTGAACGCAACCAATGATAAAGTAATTTAATTTGTTCAACTAACT 340
Db 280 GAATGAAGAGGAGCTGAACGCAACCAATGATAAAGTAATTTAATTTGTTCAACTAACT 339
Qy 341 TAGGAGAAATATGAACAAGATATATCGTCTCAAAATTCAGCAACCGCTGATGCTTTGG 400

```

Db 340 TAGGAGAAATATGACAAAGATATATCGTCTCAAAATCAGCAACGCCCTGAATGCTTTGG 399  
QY 401 TTGCTGTCTGAATTTGGCAGCGGGTTGTGACCAATTCACAGAAAAGGCTTCCCGCTATG 460  
Db 400 TTGCTGTCTGNAATTGGCAGCGGGTTGTGACCAATTCACAGAAAAGGCGCGGAAAC 459  
QY 461 TTACTATCTTTAGGTGAACCACTTAGCGTTAAAGCCACTTTCGGCTATGTTACTATCTT 520  
Db 460 CTGCTGCATGAAAGTGGCTCACTTAGCGTTAAAGCCACTTTCGGCTATGTTACTATCTT 519  
QY 521 TAGGTGAACACTATTCACAATCTGTTTACAGCGGGCTTACAGGAATGGATGTAG 580  
Db 520 TAGGTGAACACTATTCACAATCTGTTTACAGCGGGCTTACAGGAATGGATGTAG 579  
QY 581 TACAGCGCACAGCCACTATGCAAGTAGATGTTAAATAAACCACTTATCCGCAACAGTGTG 640  
Db 580 TACAGCGCACAGCCACTATGCAAGTAGATGTTAAATAAACCACTTATCCGCAACAGTGTG 639  
QY 641 AGCTATCATTAATTTGGAACAATTTAAACATCGACCAATGAAATGGTGGAGTTTATAC 700  
Db 640 AGCTATCATTAATTTGGAACAATTTAAACATCGACCAATGAAATGGTGGAGTTTATAC 699  
QY 701 AAGAAAACAACACTCCGCGGTATTCAACGGTTTACATCTAACCAAACTCTCCCAATTA 760  
Db 700 AAGAAAACAACACTCCGCGGTATTCAACGGTTTACATCTAACCAAACTCTCCCAATTA 759  
QY 761 AAGGATTTTATAGTTTACGCAAGTCTTTTAAATCAACCAAAATGGTATCACAATAG 820  
Db 760 AAGGATTTTATAGTTTACGCAAGTCTTTTAAATCAACCAAAATGGTATCACAATAG 819  
QY 821 GTAAGACGAATTTAATCAACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 880  
Db 820 GTAAGACGAATTTAATCAACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACG 879  
QY 881 AAAACATCAAGCGGCTAATTTACCTTTGAGCAAAACCAAGATAAAGCGCTCGCTGAAA 940  
Db 880 AAAACATCAAGCGGCTAATTTACCTTTGAGCAAAACCAAGATAAAGCGCTCGCTGAAA 939  
QY 941 TTGTGAATCAAGCTTTAATTTACTCGTTAAGACGGCAGTGAATCTTATTGTTGGCA 1000  
Db 940 TTGTGAATCAAGCTTTAATTTACTCGTTAAGACGGCAGTGAATCTTATTGTTGGCA 999  
QY 1001 AAGTGA AAAACGAGGCTGTGATTAGCGTAAATGGTGGCAGATTTCTTTACTCGCAGGGC 1060  
Db 1000 AAGTGA AAAACGAGGCTGTGATTAGCGTAAATGGTGGCAGATTTCTTTACTCGCAGGGC 1059  
QY 1061 AAAAATCACCATCAGCGATATATTAACCCAAACCAATTTACTTACAGCATTTGCGCGCTG 1120  
Db 1060 AAAAATCACCATCAGCGATATATTAACCCAAACCAATTTACTTACAGCATTTGCGCGCTG 1119  
QY 1121 AAAATGAAGCGGTCAATCTGGCGATATTTTGGCAAGCGGTAAACATTAATGTCCGTG 1180  
Db 1120 AAAATGAAGCGGTCAATCTGGCGATATTTTGGCAAGCGGTAAACATTAATGTCCGTG 1179  
QY 1181 CTGCGACTATTGCAAAACCAAGTAAACTTTCTGCTGATCTGTAAAGCAAGATAAAAGC 1240  
Db 1180 CTGCGACTATTGCAAAACCAAGTAAACTTTCTGCTGATCTGTAAAGCAAGATAAAAGC 1200  
QY 1241 GCAATATGTTCTTTCCGCAAAAGGGTGAACGGGAAATTTGGCGGTGAATTTCCGCTC 1300  
Db 1201 -----CTTTCCGCAAAAGGGTGAACGGGAAATTTGGCGGTGAATTTCCGCTC 1249  
QY 1301 AAAATCAGCAAGTAAAGCGGCAAGCTGATGATACAGGGGATAAAGTCAACATTAATAA 1360  
Db 1250 AAAATCAGCAAGTAAAGCGGCAAGCTGATGATACAGGGGATAAAGTCAACATTAATAA 1309  
QY 1361 CAGGTGACGTTATCGACCTTTACAGTAAAGAGGGGGAACAACTTACCTTGGCGGTGAG 1420  
Db 1310 CAGGTGACGTTATCGACCTTTACAGTAAAGAGGGGGAACAACTTACCTTGGCGGTGAG 1369  
QY 1421 AGCGGCGGAGGTAAAGACGCAATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCT 1480

Db 1370 AGCGCGGCGAAGGTAAAAACGGCATTTCAATTTAGCAAAAGAAACCTCTTTAGAAAAGGCT 1429  
QY 1481 CAACCATCAATGATATCAGCGCAAAAGAAAGCGGACGCGCTATTGTGTGGGCGGATATG 1540  
Db 1430 CAACCATCAATGATATCAGCGCAAAAGAAAGCGGACGCGCTATTGTGTGGGCGGATATG 1489  
QY 1541 CGTTAATTTGACGGCAATTAACGCTCAAGGTAGTGTGATATCGCTAAACCCGGTGGTT 1600  
Db 1490 CGTTAATTTGACGGCAATTAACGCTCAAGGTAGTGTGATATCGCTAAACCCGGTGGTT 1549  
QY 1601 TTGTGGAGACATTCGGGCAATTTATTCATTTGACAGCAATGCAATTTGTTAAACAAAAG 1660  
Db 1550 TTGTGGAGACATTCGGGCGATGATTTATCTCAAGACATGCAATTTGTGACGCCAAG 1609  
QY 1661 AGTGGTTGTAGACCTTGATGATGATGAACAAATTTGAAGCGAAGACCCCTTGGCAATATA 1720  
Db 1610 AGTGGTTGTAGACCTTGATGATGATGAACAAATTTGAAGCGAAGACCCGACGACGAATA 1669  
QY 1721 CCGGTATAATGATCAATTTCCACACAGGCGGCTGAAGCAAGGACCTTAACAAAATA 1780  
Db 1670 CTTCAAGACGATGAATACACGGGATCCGGAATAGTGGCAGCACCCCAAAAGCAACA 1729  
QY 1781 GCGAACTCAAAAACACGCTAACCAATACAACTATTTCAAAATTTCTGAAAAACGCTGGA 1840  
Db 1730 AAGAA ---AAGACAACATTAACAAACACAACTCTTTGAGAGTATATCTAAAAAAGGTACCT 1786  
QY 1841 CAATGAATATAGGCGATCAAGAAACTTTACCGTTAATAGCTCAATCAACATCGGAAGCA 1900  
Db 1787 TTGTTAATCACTACTCTTAATCAACGCTATGATGATCAATAGCTCAATTAATTT ---ATCCA 1843  
QY 1901 ACTCCACTTAATTTCTCCATAGTAAAGGTACAGCTGGCGGAGGCGTTCAGATTGATGGAG 1960  
Db 1844 ATGCGACTTAATCTTTGGAGTGGGCTCGCGGCGTGGCGGCTTGAGATTACAAACG 1903  
QY 1961 ATATTAG-----TTCTAAAGCGGAAATTTAACCATTTATCTGCGGATGGG 2008  
Db 1904 ATATTACCCGCTGATGATACAGAGGTGCAAACTTAACAAATTTACTCAGCGCTGGG 1963  
QY 2009 TTGATGTTCAATAAATAATTAACGCTTGATCAGGCTTTTAAATATATACCGCGCTCCG 2068  
Db 1964 TTGATGTTCAATAAATAATTAACGCTTGATCAGGCGCAAGGTACATTAACATTTACAGCTAAC 2023  
QY 2069 TAGCTTTGAGGTGAATTAACAAAGCAGCGCGGCAATGCTAAATTTGTCGCC 2128  
Db 2024 AAGATATCGCTTTGAGAAAGGAAGCAACCAAGTC-----ATTACAGGTC 2068  
QY 2129 AGGCACTGTAACTTACAGAGAGGAAAGATTTCCAGGCTTAACACGATATCTTTAA 2188  
Db 2059 AAGGACTATTACC---TCAGGCAATCAAAAAGGTTTTAGATTTTAATGATGCTCTCTAA 2125  
QY 2189 ACGGAACGGGTAAAGCTCTGAATATCATTTTCATCATGTAATTAATTTA-----ACCC 2239  
Db 2126 ACGCACTGCGACGCGACTGCAATTCACCACATAAAGAACCAATAATACGCTATCAAA 2185  
QY 2240 ACAATCTTAGTGCAATTAACATATCTGGGAATTAACAAATTAACCAACTA---CGA 2296  
Db 2186 ATAAATTTGAAGGACTTTAAATATTTTCAAGGAAAGTGAACATCTCAATGGTTTTACCTA 2245  
QY 2297 GAAAGAACCTCTGATTTGGCAACACGCGCATGTTTCGCTACGTGCAAGCTGCTCTTA 2356  
Db 2246 AAAATGAAGTGGATATGATTAATTCAGGACGCACTTACTGGAATTTAACTCGAAG 2305  
QY 2357 ATCTAGAGACGCGCAATTTTACCCT---TATTAAATACATTTCAAGCAATAGCAAG 2413  
Db 2306 TGGATATGATAATTCAGGACGCGCTCACTATTGACTCCAGAGGAGCGATGATGCGAG 2365  
QY 2414 GCTTAACACACAGTATAGAGCTCTCAGGCGTGAATTTTAAAGCGGTAAATGCGACAA 2473  
Db 2366 GCACACTTACCAGCGCTTATTAATTTAAACGGTATATCATTCACCAAGACACT---ACCT 2422  
QY 2474 TGTCAATTCATCTCAAGAGGCGGAAAGTTAATTTCAAATTTAAACCAACCAAGCAACA 2533  
Db 2423 TTAATGTTGACGAATCAAGAGTCAACTTTGACATCAAGCACAATAGGATAAATA 2482



Db 4628 TCAATACATTCACCGGTATAGCAAGCGTACATGAAGTAATGAAGCGAAGCATCC 4687  
 QY 4559 TTGAAAAGTAAAGATTTATCTGATGAAGAAAGAGAAACATTAAGTAACTTGGTGTA 4618  
 Db 4688 TTGAGAAGGTAAGATTTATCTGATGAAGAAAGAGAAAGCGCTAGCTAAACTTGGCGTAA 4747  
 QY 4619 GTGCTGACGTTTGTGAGCCAAATATACAAATACAGCTCAATACACAAATGAATTTA 4678  
 Db 4748 GTGCTGACGTTTGTGAGCCAAATATACAAATACAGCTCAATACACAAATGAATTTG 4807  
 QY 4679 CAACACAGCGTCAAGTCAAGTGATAATTTCTCAAGGTAAGCGGTCTTCTCAAGTGTA 4738  
 Db 4808 CAACACAGCGTCAAGTCAAGTGATAATTTCTCAAGGTAAGCGGTCTTCTCAACACAGT 4867  
 QY 4739 ATGGCGCAGCAGTATGACCAATGTGCTGACGATGGACGCGGTAGTCAAGTAATGACA 4798  
 Db 4868 ATGGCGCAGCAGTATGACCAATGTGCTGACGATGGACGCGGTAGTCAAGTAATGACA 4927  
 QY 4799 AGGTAGATTTCACTCTGCAATCAAGTCAATTTATTTCTGATATTTACTGTGGGTTA 4858  
 Db 4928 AGGTAGATTTCACTCTGCAATCAAGTCAATTTATTTCTGATATTTACTGTGGGTTA 4987  
 QY 4859 AAGTTCAGTACGGCTTTACCCATCTTGTAAAAAATACGGAGAATACAAATGAATTT 4918  
 Db 4988 AAGTTCAGTACGGCTTTACCCATCTTGTAAAAAATACGGAGAATACAAATGAATTT 5047  
 QY 4919 TTAACAGGTTATTTATG 4937  
 Db 5048 TTAACAGGTTATTTATG 5066

## RESULT 14

AAT90993

ID AAT90993 standard; DNA; 4803 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

CC 2 Novel nucleic acids (AAT90992 and AAT90993) respectively comprise the  
 CC coding sequences of high molecular weight surface proteins HMW3  
 CC (see AAW30291) and HMW4 (see AAW30292) of Haemophilus influenzae strain  
 CC 5. HMW3 has a mol. wt. of 125 kDa and HMW4 has a mol. wt. of 123  
 CC kDa. Both are antigenically related to HMW1 (see AAW30293) and HMW2  
 CC (see AAW30294) obtained from non-typeable H. influenzae strain 12.  
 CC Vectors containing HMW nucleic acids for transformation of a host  
 CC are claimed. HMW nucleic acids can also be used as probes for the  
 CC detection of related genes. HMW proteins, conjugates and peptides  
 CC can be used in vaccines, as immunogens for preparation of  
 CC antibodies and as antigens for detection of these antibodies.  
 XX

SQ Sequence 4803 BP; 1714 A; 928 C; 971 G; 1190 T; 0 other;

Query Match 45.5%; Score 2245; DB 18; Length 4803;  
 Best Local Similarity 70.6%; Pred. No. 0;  
 Matches 3411; Conservative 0; Mismatches 990; Indels 432; Gaps 14;

QY 352 ATGACCAAGATATATCGTCTCAAAATTCAGCAACGCGTGAATGCTTGGTGGTGTCT 411  
 Db 1 ATGACCAAGATATATCGTCTCAAAATTCAGCAACGCGTGAATGCTTGGTGGTGTCT 60  
 QY 412 CAATTGGCAGGGGTTGTGACCATTCACAGAAAAAGGCTTCCGCTATGTTACTCTTT 471  
 Db 61 GAATTGACAGGGGTTGTGACCATTCACAGAAAAAGGCTGAAAAACCTGTCGTACG 120  
 QY 472 AGGTGTAACACACTAGCGTTAAAGCCACTTTCCGCTATGTTACTCTTTAGCTGTA 531  
 Db 121 AAGTACGCCACTTGGCGTTAAAGCCACTTTCCGCTATGTTACTCTTTAGCTGTA 180  
 QY 532 TCTATTCACAACTCTGTTTTAGCAAGCGGCTTACAGGAATGAGTGTACACGCGACA 591  
 Db 181 TCCATTCCGCAATCTGTTTTAGCGAGGGTTTACAGGGAATGAGCGTCTGACGCTACA 240  
 QY 592 GCCACTATGCACTAGTGTGTAATAAACCCATTATCCGCAACAGCTGTACGCTATCAT 651  
 Db 241 GCACCATGCACTAGTGTGTAATAAACCCATTATCCGCAACAGCTGTACGCTATCAT 300  
 QY 652 AATTGGAACAATTTAAACATCGACCAAAATGAAATGGTGCAGTGTTCACAGAAAAAC 711  
 Db 301 AATTGGAACAATTTAAACATCGACCAAAATGAAATGGTGCAGTGTTCACAGAAAAAC 360  
 QY 712 AACTCCCGCTTATCAACCGTTTACATCTACCAAAATCTCCCAATTAATAAGGGATTTA 771  
 Db 361 AACTCCCGCTTATCAACCGTTTACATCTACCAAAATCTCCCAATTAATAAGGGATTTA 420  
 QY 772 GATTCTAACGCAAGTCTTTTTTAACACCAAAATGATATCAATAGTAAAGACGCA 831  
 Db 421 GATTCTAACGCAAGTCTTTTTTAACACCAAAATGATATCAATAGTAAAGACGCA 480  
 QY 832 ATTTAAACACTAATGGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAAACATCA 891  
 Db 481 ATTTAAACACTAATGGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAAACATCA 540  
 QY 892 GCGGTAAATTTCACTTCAGCAAAACCAAGATGAAAGCGTCTCGTGAATTTGTAATCAC 951  
 Db 541 GCGGTAAATTTCACTTCAGCAAAACCAAGATGAAAGCGTCTCGTGAATTTGTAATCAC 600  
 QY 952 GGTTTAATTTACTCTGCTTAAAGAGCGGCTGTAATTTCTTATGTCGCAAGTGAATAAC 1011  
 Db 601 GGTTTAATTTACTCTGCTTAAAGAGCGGCTGTAATTTCTTATGTCGCAAGTGAATAAC 660  
 QY 1012 GAGGCTGATTTAGGCTAAATGGTGGCAGCATTTCTTTACTCGCAGGCAAAAAATCAC 1071  
 Db 661 GAGGCTGATTTAGGCTAAATGGTGGCAGCATTTCTTTACTCGCAGGCAAAAAATCAC 720  
 QY 1072 ATCAGCGATATAATAAACCCCAACCATTTACTTACAGCATTTGCGCGCTCGCTGAAATGA 1131  
 Db 721 ATCAGCGATATAATAAACCCCAACCATTTACTTACAGCATTTGCGCGCTCGCTGAAATGA 780  
 QY 1132 GTCATCTGGCGATATTTTTCGCAAGGCGGTAAACATTTATGTCGCTGCCACTATT 1191  
 Db 781 ATCAATCTGGCGATATTTTTCGCAAGGCGGTAAACATTTATGTCGCTGCCACTATT 840

14-APR-1998 (first entry)

Non-typeable Haemophilus high mol.wt. surface protein HMW4 gene.

Non-typeable Haemophilus; high molecular weight surface protein;

HMW4; immunogen; vaccine; otitis media; ss.

Haemophilus influenzae strain 5.

Key Location/Qualifiers

CDS 1..4803

/tag= a

/trans\_except= (pos:1114..1116, aa:Thr)

/trans\_except= (pos:1198..1200, aa:Asp)

W09736914-A1.

09-OCT-1997.

01-APR-1997; 97MO-US04707.

01-APR-1996; 96US-0617697.

(BARE/) BARENKAMP S J.

Barenkamp SJ;

WPI; 1997-503038/46.

P-PSDB; AAW30292.

High molecular weight proteins of non-typeable Haemophilus

influenzae - useful for vaccine production

Claim 1; Page 90-92; 183pp; English.



QY	1192	CGAACCAAGGTAACATTTCTGCTGATCTCTGTAAGCAAGATATAAAGCGCAATATGTT	1251	QY	2266	TCGTGGAATATAACATTTAACCAAACTACGAGAAAGAACACCTCGTATTTGGCAACACGAC	2325
Db	841	CGCAATAAAGGTAAATTTCTGCGCACTCTGTAAGCAAGATATAAAGGTGTAACATTTGTT	900	Db	1906	TCGTGGAATATAACATTTAACCAAACTACGAGAAAGAACACCTCGTATTTGGCAACACGAC	1965
QY	1252	CTTTCCGCAAGAGGTTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAA	1311	QY	2326	CATGATTCGCACTGGAAGCTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTT	2385
Db	901	CTCTCTGCCAAGAGGTTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAA	960	Db	1966	AAAGACTCTCTCTGGAATGTTCTCTCTTACTTTGTAATCGGTGCAAAATTTTACCTTT	2025
QY	1312	GCTAAAGCGCAAGCTGATGATTTACAGGGGATAAAGTACATATAAACAACAGGTGCAAT	1371	QY	2386	ATTAATATACATTTTCAAGC-----AATAGCAAGGCTTAACACACACAGTATAGAGCTCT	2439
Db	961	GCAAAAGTGGTAAGTTGATGATTTACAGGTGATAAAGTACATATAAACAACAGGTGCAAT	1020	Db	2026	ATAAATTCGTTGATAGCGCTCAATTTCCCAAGATTTTGGGTGATCAGCTAGAGGTTTT	2085
QY	1372	ATCGACCTTCAGGTAAGAGGGGAGAACTTTACCTTGGCGGTGACGAGCGCGCGAA	1431	QY	2440	GCAGGGGTGAATTTTAAACGCGTAATGCAACATGTCATTAATCTCAAGAGAGGAGCG	2499
Db	1021	ATCGACCTTCAGGTAAGAGGGGAGAACTTTACCTTGGCGGTGATGAGCGTGGCGAA	1080	Db	2086	GCAGGGGTGAATTTTAAACGCGTAATGCAACATGTCATTAATCTCAAGAGAGGAGCG	2145
QY	1432	GTTAAACCGCATCAATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAAT	1491	QY	2500	AAAGTTAAATTTCAAAATTAACCAACAGAGACATGAACACAAAGCAAACTTTTACCAAT	2559
Db	1081	GGTAAATATGTTATTCATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAAT	1140	Db	2146	AAAGCTTTAATTTAAATTTAAACCAACAGAGCGCGCTACAGACCCCAAAAGAAATTTACCTAT	2205
QY	1492	GTATCAGGCAAGAAAGGCGGCGCTATTGTTGGCGGATATTGCGGTTAAATTTGAC	1551	QY	2560	CGGTTTTTAGCCTAATATCACAGCAGCTG-----GTGGGCTCTGTTTTTTTGGATATA	2613
Db	1141	GTATCAGGCAAGAAAGGCGGCGCTATTGTTGGCGGATATTGCGGTTAAATTTGAC	1200	Db	2206	ACTTTTAAAGCGCAACATTTACAGCTACCGGTACCGGTAAACAGTATGCTGTTGACATA	2265
QY	1552	GGCAATATTAAACGTCAGGTTAGTATGATCGCTAAACCGGTGTTTGTGGAGACA	1611	QY	2614	TATGCAACCACTTCTGGCAGAGGGGCTGATTTAAATGAGTGAATTAATATCTTAAC	2673
Db	1201	GGTAAATATGTTATTCATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAAT	1257	Db	2266	CACGCCAATCTTACCTCTAGAGCTGCGCGCATTAACATGGAATTTCAATTAACATTACCGC	2325
QY	1612	TCGGGGCATTTATTTATCCATGACAGCATGCAATGTTTAAACAAAGAGTGTGCTA	1671	QY	2674	GGCGCTAATTTTACCTTAAATTTCCCATTTTGGGGGCTGATGACGCTTTTAAATCAACAA	2733
Db	1258	TCAGGACATGATCTATCCATTTGATGATGTTGTTGTCAGCTTAAAGAGTGTATTA	1317	Db	2326	GGCTTGATTTTCCCATTAACATCCCATTAATGCGCAATAGTATGCTTTGAAATCAAAA	2385
QY	1672	GACCTGATGATGTAACAAATTTGAAGCGGAAGACCCCTTGGCAATTAATCCGCTATAAT	1731	QY	2734	GACTTAACCATTAATGCAACCAATTTCAAATTTTCCGCTCAGACAGAGCAAGAGTATTT	2793
Db	1318	GACCCAGATGATGTTGTCATTTGAAACTCTTACATCTGGAGCAATTAATCCGCGAAG	1377	Db	2386	GACTTAACCATTAATGCAACCAATTTCAAATTTTCCGCTCAGACAGAGCAAGAGTATTT	2445
QY	1732	GATGAATTTCCCAACAGGCGGCTGAAGCAAGCGACCTTAAACAAATAGCAACCTCAAA	1791	QY	2794	TATGAGGCTGACGAGCAATGCGCAATGCGCATCAATCAACCTTCAACATATCCATTTGGGGGT	2853
Db	1378	CAAGGATATACCAACAGGAGATGGGACTTAAAGAGTCACTTAAGGTATAGTATTTCTAA	1437	Db	2446	TATGAGGCTGACGAGCAATGCGCAATGCGCATCAATCAACCTTCAACATATCCATTTGGGGGT	2505
QY	1792	ACAACGCTAACCAATACAACTATTTCAAATTTATCTGAAACGCGCTGGCAATGAATATA	1851	QY	2854	ATGTCACCTCTGCTGCGCAACCAAACTCAAGCAGCAGCATTTACGGGGAATTTACTACTAG	2913
Db	1438	CCTACATTAACAACTCAACTCTTGAGCAAAATCTTGAAGAGGTTCTTTATGTTAATATC	1497	Db	2506	ATGTCACCTCTGCTGCGCAACCAAACTCAAGCAGCAGCATTTACGGGGAATTTACTACTAG	2565
QY	1852	ACGCGATCAAGAAACTTTACCGTTAATAGCTCAATCAACATCGGAACCACTCCACTTA	1911	QY	2914	AAAGCAGCAATGTTACGCTTAGAAGCAATTAACGCCCTTAATCAGCAAAACATAAGGAT	2973
Db	1498	ACTGCTAATAAGATTTATGTTATAGCTCCATCAACTTT---ATCTATGGCAGTTTA	1554	Db	2566	AAAGCAGCAATGTTACGCTTAGAAGCAATTAACGCCCTTAATCAGCAAAACATAAGGAT	2625
QY	1912	ATTTCCCATAGTAAGGTCAGGTCAGGCGGAGGCTTCAGATTTGATGAGATATTACTTCT	1971	QY	2974	AGAGTTATAAAGCTTGGCAGCTTCTGCTTAATGCGGAGTTTAAAGTTAACTGGCGAAAT	3033
Db	1555	ACACTTCACACTAA-----CGAGATGGAGTTAAATTAAGCGGTGATATTACCTCA	1605	Db	2626	AGAGTTATAAAGCTTGGCAGCTTCTGCTTAATGCGGAGTTTAAAGTTAACTGGCGAAAT	2685
QY	1972	A---AAGCGCAATTTAACTTTATTTCTGGCGGATGGTTGATGTTCAATAAATATTT	2028	QY	3034	CGAGATATAAAGCTTGGCAGCTTCTGCTTAATGCGGAGTTTAAAGTTAACTGGCGAAAT	3093
Db	1606	ACGGAATGATGATTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAA	1665	Db	2686	CGAGATATAAAGCTTGGCAGCTTCTGCTTAATGCGGAGTTTAAAGTTAACTGGCGAAAT	2745
QY	2029	ACGCTTATGATGAGGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT	2085	QY	3094	GATACCTTAAATPATCAGCGCAATTTTACCAATTAATGCGCTACCGCAAAATTAATAAACA	3153
Db	1666	ACGCTTATGATGAGGTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTTAACTTT	1725	Db	2746	GACAACCTTAAACATCAGCGCACTTTTACCAACCAAGCTACCGCAACATTAATAAATAA	2805
QY	2086	ATAACAAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1785	QY	3154	CAAGGAGTGGTAAAACT---TGGCAATGTTTACCAATGATGCTGATTTTAAACATTTACCACT	3210
Db	1726	GGCGATTAAGCAACGTAACCAACAGATGCTCAAAATACCGCAACGAGGAGGAGGAGGAGG	1845	Db	2806	CAAGGAGTGGTAAAACT---TGGCAATGTTTACCAATGATGCTGATTTTAAACATTTACCACT	2865
QY	2146	ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2205	QY	3211	CAGCTTAAAGCGCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	3270
Db	1786	AATAAAGATGATTAACCAATTTAGATTTCAATTAATGATTTCAATTAATGATTTCAATTA	1845	Db	2866	AACGCTTCAGGCACTCAAAACCACTTATTAACGGAATATTAATTAACGGAAGGAGGAG	2925
QY	2206	CTGAATATCATTTCAATGATTAATTTAACCACATCTTAGTGCGCAATTAACATA	2265	QY	3271	TTAATATTTACAGACAGTATTAATGATGCTGAATTTTAACTTTTGGCGCAATTTCTCGCA	3330
Db	1846	TAAAGGTTTATTCGAATCAAAATTAATTTCTCTCTGATTAATTTCAACCAACAGGATTAAC	1905	Db	2926	TAAACATCAAGATATTTAAAGCGGAGGCGGCAATTTTAACTTTTGGCGCAATTTCTCGCA	2985
QY				QY	3331	AAAGAGGCAAGCTCAGGATTTTCTCTCGGATTAATTTTAACTTTTAACTTTTAACTTTTAACTTT	3390

```

Db 2986 AAAGAGGCAATCTCAGAAATTTCTTGATAAAGTAAATATACCAGTCAAGTAAATC 3045
QY 3391 AAAAGAGGATTTGATGGAGGAGCTCTAGTTGAGATGGGAGCAAGTAAATGCCAACCTAACT 3450
Db 3046 AAAGAGGCGGTTGAAGGGGGGGTCTGATTCAGTGAGGAGGAGAAATGCTAACCTAACT 3105
QY 3451 ATTAAAACCAAGAAATGAAATTCAGAGAGAGCTTAAGTATTTTCAGGTTTCAATAAAGCA 3510
Db 3106 ATTCAAAACCAAGAGTAAATTCGGCAGGAGACCTAAATATTTTCAGGCTTTAATAAAGCA 3165
QY 3511 GAGATTACAGCCAAAGATGGTAGAGATTTAACTATTGGCAACAGTAAATGACGGTAAACAGC 3570
Db 3166 GAATATTACAGCTAAATAATTCGGCAGTGAATTTAACTATTGGCAATGCTAGCGGTGGTAA 3221
QY 3571 GTGCCGAGCCAAACAGTAACTTTTAACTATGTTAACTATGTTAACTATGTTAACTATGTTAA 3221
Db 3222 --TGCTGATGCTTAAAGAGTGAATTTTGAACAGTAAATGTTAACTATGTTAACTATGTTAA 3630
QY 3631 GGTCAATGTGACACTAAATAGCAAGTGAATTTTGAACAGTAAATGTTAACTATGTTAACTAT 3279
Db 3280 GTTCAATGTGACACTAAATAGCAAGTGAATTTTGAACAGTAAATGTTAACTATGTTAACTAT 3690
QY 3691 AGCAATTAGCAAGTAACTATGTTAACTATGTTAACTATGTTAACTATGTTAACTATGTTAA 3333
Db 3334 GCTGATTAATGATACAGCAGCAGCTTTAACTATGTTAACTATGTTAACTATGTTAACTATG 3750
QY 3751 GATATTACTTCTCTCAAAACAGTAAATATGTTAACTATGTTAACTATGTTAACTATGTTAA 3393
Db 3394 ATCGTTACTTCCCAAGAGCAATAAATATCTCTGCGCGCAGGAAATGTTAACTATGTTAA 3798
QY 3799 ----- 3798
Db 3454 GAAGGCACAATATCAATGCAACACAGGAGCGTGGAGTAACTGTTCAAAATGGTACA 3513
QY 3799 ----- 3798
Db 3514 ATTAAGGCAACATTACCTGCAAAATGTAACAGTGTACAGCAACAGAAATCTGTTAAC 3573
QY 3799 ----- 3798
Db 3574 ACAGAGAAATGCTGTCATTAAATGCAACACAGGAGCGTGGAGTAACTGTTCAAAATGGTACA 3633
QY 3799 ----- 3798
Db 3634 GATATTAAAGGTGGAATTAATCACTTCCGGTAAATGTAATATTACAGCGAGCGCAAT 3693
QY 3799 ----- 3798
Db 3694 ACATTAAAGGTAACTAATATCACTGTTCAAGATGTAAACAGTAAACAGCGGATGCAAGGCC 3753
QY 3799 ----- ACCACACAGCAGCTCGACCATTAACGCAACAAATGGCAAGCAAGTATTACAGC 3855
Db 3754 TTGACAACTAGCAGCGCTCAACCATTAAGTGGCAACAGCGCAATGCAAAATATTACAGC 3813
QY 3856 AAAACAGGTGATATCAGC----- 3873
Db 3814 AAAACAGGTGATATCAACGGTAAAGTGAATCAGCTCCGGCTCTGTAACTGTTGCA 3873
QY 3874 -----GGTACGATTTCCGGTAAACAGCGTAAAGTGTAGCGC--- 3909
Db 3874 ACTGGACCACTCTGCTGTAGTAAATATTTCAGGTAACACTGTTACTATTACTCGGAT 3933
QY 3910 ACTGGTATTTAACTAAATCCGGCTCAAAATTTGAGCGAAATCGGGTGAAGCTAAT 3969
Db 3934 AGCGGTAATTAATTAACCTCCACAGTAGGTCTACAAATTAATGGGACTAATAGTGTAAACCA 3991
QY 3970 GTAAAGTGAACAGTACAAATGCGGTACAAATTTCCGGTAAATACGTTAAATGTTACG 4029
Db 3992 -CCTCAAGCAATCAGCGATATTGAAGGTACAATTTCTGGTAAATACAGTAAATGTTACA 4050
QY 4030 GCAACGCTGGCGATTTAAACAGTGGGAATGGCGCAAAATTAATGCCACAGAGAGCT 4089

```

```

Db 4051 GCAAGCACTGGTGATTTAACTATTGAAATAGTGCAAAAGTTGAAGCGAAAATGAGCT 4110
QY 4090 GCAACCTTAACCCCAACAGGAATACCTTGACTACTGAAGCGGTTCTAGCATCACTTCA 4149
Db 4111 GCAACCTTAACCTGCTGAATCAGGCAAAATTAACCCCAACAGGCTCTAGCATCACTTCA 4170
QY 4150 ACTAAGGGTCAGTAGACCTCTTGGCTCAGAAATGTCAGATGTCGAGAAAGCAATTAAGCT 4209
Db 4171 AGCAATGGTCACAACTCTTACAGCAAGGATAGCATATCGCAGAAACAAATTAAGCT 4230
QY 4210 GCTAATGTCACAACTCTTACAGCAAGGATAGCATATCGCAGAAACAAATTAAGCT 4269
Db 4231 GCTAATGTCAGCACTCTTACAGCAAGGATAGCATATCGCAGAAACAAATTAAGCT 4290
QY 4270 GCAACAGCGCACTCTTGGTATTAAACGCAAAAGATGTCAGTAAATGGTGAATGATCA 4329
Db 4291 GCAACAGCGCACTCTTGGTATTAAACGCAAAAGATGTCAGTAAATGGTGAATGATCA 4350
QY 4330 GGTGATAGTACAGAGTGAATGTCAGTAAACGCAAAAGATGTCAGTAAATGGTGAATGATCA 4389
Db 4351 GGTGATAGTACAGAGTGAATGTCAGTAAACGCAAAAGATGTCAGTAAATGGTGAATGATCA 4410
QY 4390 ACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGTGAATGATCA 4449
Db 4411 ACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGTGAATGATCA 4470
QY 4450 TCGAAAGATGTTAGAAACACTGTCGCTTAAAGGCAAGGAAATTTAGTGAATGATCA 4509
Db 4471 TCGAAAGATGTTAGAAACACTGTCGCTTAAAGGCAAGGAAATTTAGTGAATGATCA 4530
QY 4510 CAGCAGGCTGAGCAAGTGTAGAAGAGTAAATGAAGCGAAACGCTCTCTGAAAGGTA 4569
Db 4531 CAACAGGCTGAGCAAGTGTAGAAGAGTAAATGAAGCGAAACGCTCTCTGAGAGGTA 4590
QY 4570 AAAGATTTATCTGATGAAGAGAAACATTTAGCTAACTTGGTGAAGTGAATGATCA 4629
Db 4591 AAAGATTTATCTGATGAAGAGAAACATTTAGCTAACTTGGTGAAGTGAATGATCA 4650
QY 4630 TTTGTTGAGCCAAATTAATACAAATTTACAGTCAATACAAATTAATTTACACAGGAG 4689
Db 4651 TTTGTTGAGCCAAATTAATACAAATTTACAGTCAATACAAATTAATTTACACAGGAG 4710
QY 4690 TCAAGTCAAGTGAATTTCTGAGAGTGAAGCGGTTTCTCAAGTGAATGGCGCACCA 4749
Db 4711 TCAAGTCAAGTGAATTTCTGAGAGTGAAGCGGTTTCTCAAGTGAATGGCGCACCA 4770
QY 4750 GATATGACCAATGTTGCTGACGATGACAGCGG 4782
Db 4771 GATATGACCAATGTTGCTGACGATGACAGCGG 4803

```

## RESULT 15

AA049511

ID AA049511 standard; DNA; 4702 BP.

XX AA049511;

XX 26-APR-1994 (first entry)

XX Partial sequence encoding high molecular weight protein 4 (HMM4).

XX HW; high molecular weight protein; virus; vaccine; Influenza;  
 KW epitope; immunity; haemophilus influenzae; ss.  
 XX Haemophilus Influenzae.

OS Key Location/Qualifiers

FH CDS 2..4591

FT /\*tag= a

FT /\*product= High molecular weight protein 4.

FT /\*tag= b

FT /\*transl\_except= TCT encodes Threonine.

FT misc\_difference 899..901

FT

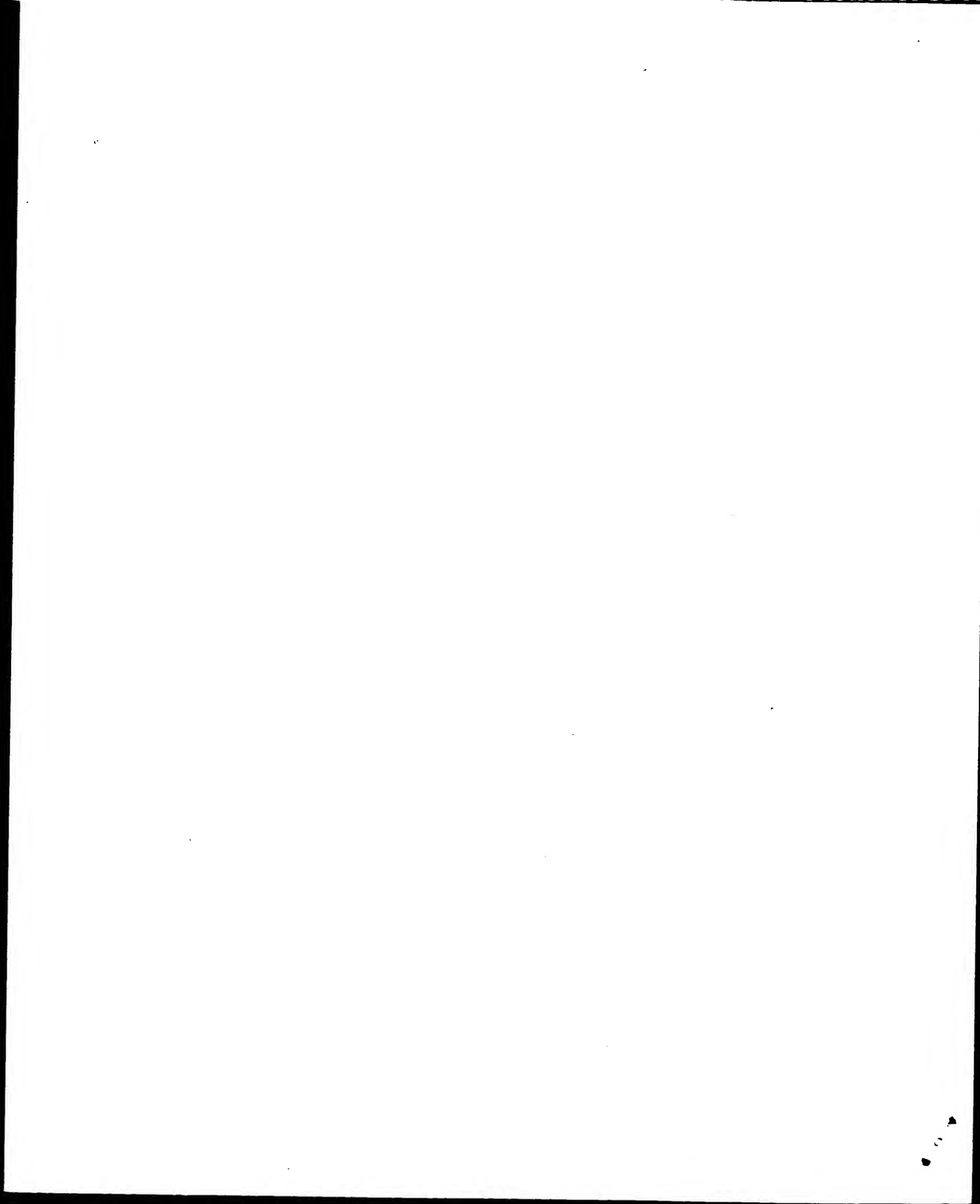


1607 TCTATTAAACGGGAGCGGAGGTTTAAAGTTTATTGCAAAATCAAAATTAATTTCACTCAT 1666  
2242 AATCTTAGTGGCACAATTAACATATCTGGGAATATTAACAANTTAACCAACTACGAGAAG 2301  
1667 AAATTTGATGGCGAAATTAACATATCTGGGAATATTAACAANTTAACCAACTACGAGAAG 1726  
2302 AACACCTCGTATTGGCAACAGCCATGATTGCGACTGGAAGCTGAGTCTCTTAATCTA 2361  
1727 GATGTTAAATCTGGNATGCAATCAAAAGACTCTTACTGGAATGTTTCTTCTTACTTTG 1786  
2362 GAGACAGGCGCAATTTTACCTTTTAAATACATTTCAAGC-----AATAGCAAGGC 2415  
1787 AATACGGTGCAGAAATTTACCTTTTATAAATTCGTTGATAGCGGCTCAAAATTCGCCAAGT 1846  
2416 TTACAACACAGTATAGAGCTCTGAGGGGTGAATTTTACGGCGTAAATGGAACATG 2475  
1847 TTGAGGTCATCAGCTAGAGTTTTCAGGGGTACATTTTAAAGGTCATCGGAGCAAAACA 1906  
2476 TCATTCATCTCAAGAGGAGCGAAAGTTAATTTCAATTTAAACCAACAGAGACATG 2535  
1907 AACTTCAACATCGAGCTAACGCCAAAGCCTTATTTAAATTTAAACCAACAGCGCTACA 1966  
2536 RACACAGCAACCTTTACCAATTCGGTTTTAGCCAATATCACAGCCACTG-----GT 2589  
1967 GACCCAAAGAAATTTACCTATTACTTTTAAAGCCCAATTTACAGCTACCGGTAAACAGT 2026  
2590 GGGGCTCTGTTTTTATATATATGCAACCATTTCTGGCAGAGGGCTGAGTTAAAA 2649  
2027 GATAGCTCTGATGTTGACATACAGCCCATCTTACCTAGAGCTGCCGGCATAAAC 2086  
2650 ATGAGTGAATTAATATCTTAACGGGGCTAATTTTACCTTTAAATTTCCCATGTTCCGGGC 2709  
2087 ATGGATTCAATTAACATTTACGGGGGCTTGACTTTTCCATTAACATCCCATTAATCGCAAT 2146  
2710 GATGACCTTTTAAATCAACAAAGCTTAACCATTAATTAATTAACCAATTTCAATTTGAC 2769  
2147 AGTAATCTGTTGAATCAAAAGACTTTACTATTAATTAATTAATTAATTAATTTAGT 2206  
2770 CTCAGACAGCAAGATGATTTTATGAGGGTACGCAACCAATGCCATCAATTAATCAAC 2829  
2207 CTTAAGCAACAGAGATCTTTTATATGAATACAGCAACAGCCATTAATCTCAAGT 2266  
2830 TACACATATCCATTTCTGGGGGTAAATGTCACCCCTGTTGGCAACAACTCAAGCAGCAGC 2889  
2267 CATATCTAACCATTTCTGGGGCAATGTCACCTAGTGTGGGGAATTTCAAGCAGTAGC 2326  
2890 ATTACGGGGAATTTACTATCGAAGAGCAGCAATTTTACGCTAGAGCCCAATTAACGCC 2949  
2327 ATTACGGGCAATATCAATATCAACATTAAGCAATTTTACATTAACAGCTGACACACAGC 2386  
2950 CCAATCAGCAACATTAAGGGATAGAGTTTATAAATTTGGCAGCTTGTCTGTTAATGGG 3009  
2387 AACAGCAACAGGGCTTGAAGAAAGAACTTAACCTTTGGCAATATATCTGTTGAGGG 2446  
3010 AGTTTAAGTTTAACTGGGCAATGAGATATTAAGGCAATCTCACTATTTCAAGCAAGC 3069  
2447 AATTAAGCTAATCTGTTGCAATTTGCAACATTTGCGGCAATCTTCTTATTGCAAGAT 2506  
3070 GCCATTTTAAAGGAAGACTAGAGTACCCTTAAATATACCGGCAATTTTACCAATTAAT 3129  
2507 TCCACATTTTAAAGGAAGCCAGTGACAACCTTAACATATCACCGGCACTTTTACCAACAC 2566  
3130 GGCACCTCCCAATTAATTAACACAGAGGAGTGGTAAACCT---TGGCAATTTTACCAT 3186  
2567 GGTACCGCCCAACATTAATTAACAAAGAGGTGGTAAACCTTCCAGGCGATATTATCAAT 2626  
3187 GATGCTGATTTAAACATTTACCACTACCGCTTAAACGCAACCAAGAGCATCATCGCGGA 3246  
2627 AAGGCTGGTTTAAATATCACTACTACGCGCTCAGGCACTCAAAACCAATTTAATACGGA 2686  
3247 GATTAATCAACAAAGAGGAGCTTAAATATATACAGACAGTAAATATGATGCTGAAATC 3306

2687 AATATAACTAACGAAAAAGGCGACTTAAACATCAAGAATATTAAGCCGACGCCGAAATC 2746  
3307 CAATTTGGCGGCAATATCTCGCAAAAGAGCAACCTCAGGATTTCTTCGATAAATTT 3366  
2747 CAATTTGGCGGCAATATCTCACAAGAGAGCAATCTCAAAATTTCTTCTGATAAAGTA 2806  
3367 AATATCACCAACACAGATAACAAATCAAAAGGGTATTGATGAGAGGACTCTAGTTTCAGT 3426  
2807 AATATTACCAATCAGATACAAATCAAAAGCGGCTTGAAGGGGGGCTTGTGATTCAGT 2866  
3427 GCGACAAGTAATGCCAAGCTTAACCTTATTAACCAAGAAATTTGAATTTGACAGAGACCTA 3486  
2867 GAGCAAGAAATGCTTAACCTTAACCTTATTAACCAAGAGTAAATTTGCGAGGAGACCTA 2926  
3487 AGTATTTCAGGTTTCAATTAAGCAGAGATTACAGCCCAAGATGATGATGATTTAACTATT 3546  
2927 AATATTTCAGGCTTTAATTAAGCAGAAATTTACGCTAAATTTGCGAGTGTATTACTATT 2986  
3547 GGCACAAGTAATGACGGTAAACGGGTGCCGAAGCCAAACAGTAATTTTAAACAAATGTT 3606  
2987 GGCAATGCTAGCGGTGGTAA-----TGCTGATGCTAAAAAGTGACTTTTGACAAGGTT 3040  
3607 AAGATTTCAAAATCTCTGCTGAGCGCTACAAATGTGACACTTAATTAAGCAAGTGAACACA 3666  
3041 AAGATTTCAAAATCTGCACTGAGCGGTCAAAATGTAACTTAATTAAGCAAGTGAACAGC 3100  
3667 TCTAGCAGCAATGGCGGACCTGAAAGCAATAGCACAACGATACCGGCTTAATTAATTTACT 3726  
3101 TCT-----AATGGTAGTAGCAATGCTGGTAAATGATTAACAGCACCGGTTTAAACATTTCC 3154  
3727 GCAAAATTTGTAAGATTAACAAAGATATTTACTTCTCTCAAAACAGTAATTAATTAACCGC- 3785  
3155 GCAAAAGATGTAAACGGTAAACAAATTAACGTTTACCTCCCAAGACAAATAATTTCTGCC 3214  
3786 ----- 3785  
3215 GCAGCAGGAATGTAAACCAACCAAGAGCACAACACTATCAATGCAACACAGGCGAGTG 3274  
3786 ----- 3785  
3275 GAAGTAATGCTCAAAATGGTACAAATTAAGGCAACATTACCTCGCAAAATGTAACAGTG 3334  
3786 ----- 3785  
3335 ACAGCAACAGAAATCTTTGTACCACAGAGAATGCTGTCAATTAATGCAACACAGGCGACA 3394  
3786 ----- 3785  
3395 GTAAACATTTAGTACAAAACAGGGGATATTAAAGGTGGAATTTGAATCACTTCCGGTAAT 3454  
3786 ----- 3785  
3455 GTAAATATTACAGCGAGCGCAATACACTTAAGGTAAATTAATCACTGGTCAAGATGTA 3514  
3786 -----GTCCGAAAGGTTTACCACACAGCAGGCTCGACCAATTAACGCAACA 3831  
3515 ACAGTAACAGCGGATGAGGAGGCTTGACAACACTACAGCAGGCTCAACCAATTAAGTCCGACA 3574  
3832 AATGGCAAGCAAGTATTTACACCAACCAAGAGTGAATATCAGCGGTACG----- 3879  
3575 ACAGGCAATGCAAAATTTTACACCAACCAAGAGTGAATATCAACGCTAAAGTTGAATCCAGC 3634  
3880 -----ATTTCCGGT 3888  
3635 TCGGGCTCTGTAACTTTGTAACCTGGGCAACACTTTGCTGTAGGTAAATTTTCAGGT 3694  
3889 AACACGGTAAGTGTAGCGGACT---GGTGAATTTAACCACTAAATTCGGGCTCAAAATTT 3945  
3695 AACACTGTACTATTACTTCGGGATAGCGGTAAATTTAACTTCCACAGTACGTTCTTACAAAT 3754  
3946 GAAGCGAATTCGGGTGAGGCTTAATGTAACAACAGTGCACAGTCAATTTGGCGGTACAAAT 4005  
3755 AATGGGACTAATAGTGTAAACCA---CCTCAAGCAACATCAGCGGATATTGAAGTACAAAT 3811













```

QY 4441 AATATCATTTTCGAAGATGTAAGAACACATGTCGCTTAAGAGGCAAGAAATGAGGTG 4500
Db 4441 AATATCATTTTCGAAGATGTAAGAACACATGTCGCTTAAGAGGCAAGAAATGAGGTG 4500
QY 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAAGAAGTAATTTGAAGGCAACCGTCCTT 4560
Db 4501 AATATATCCAGCAGGTGTAGCAAGTGTAGAAGAAGTAATTTGAAGGCAACCGTCCTT 4560
QY 4561 GAAAGAGTAAAGATTTATCTGTATGAAGAAGAGAAACATTTAGCTAAACTTGGTGAAGT 4620
Db 4561 GAAAGAGTAAAGATTTATCTGTATGAAGAAGAGAAACATTTAGCTAAACTTGGTGAAGT 4620
QY 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
Db 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTTACAGTCAATACACAAATGAATTTACA 4680
QY 4681 ACCAGCCGTCAGTCAAGTCAAGTCAATTTCTGGAAGGTAAAGCGGTCTTCTCAAGTGAAT 4740
Db 4681 ACCAGCCGTCAGTCAAGTCAAGTCAATTTCTGGAAGGTAAAGCGGTCTTCTCAAGTGAAT 4740
QY 4741 GCGCCAGCAGTATGTACCAATGTTGCTGACGATGACAGCGGTAGTCAGTAATTCACAAG 4800
Db 4741 GCGCCAGCAGTATGTACCAATGTTGCTGACGATGACAGCGGTAGTCAGTAATTCACAAG 4800
QY 4801 GTAGATTTTCATCTGCAATGAAGTCATTTTATTTTCGTAATTTACTGTGGGTTAAA 4860
Db 4801 GTAGATTTTCATCTGCAATGAAGTCATTTTATTTTCGTAATTTACTGTGGGTTAAA 4860
QY 4861 GTTCAGTACGGGCTTACCCTATCTTGTAAATAATTTACGAGATACAAATAAAGTATTTT 4920
Db 4861 GTTCAGTACGGGCTTACCCTATCTTGTAAATAATTTACGAGATACAAATAAAGTATTTT 4920
QY 4921 AACAGGTTATTATTATG 4937
Db 4921 AACAGGTTATTATTATG 4937

```

## RESULT 2

```

US-08-302-832-3
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302.832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22.651

```

```

; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-3

```

```

Query Match 100.0%; Score 4937; DB 1; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TAAATATACAGATATATAAATAATCAAGATTTTGTGATGACAAACAACAATTAACA 60
Db 1 TAAATATACAGATATATAAATAATCAAGATTTTGTGATGACAAACAACAATTAACA 60
QY 61 CACCTTTTGTGAGTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATA 120
Db 61 CACCTTTTGTGAGTCTATATGCAAAATATTTTAAAAAATAGTATAAATCCGCCATATA 120
QY 121 AATGGTATAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTT 180
Db 121 AATGGTATAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTT 180
QY 181 CTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
Db 181 CTTTCATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240
QY 241 CACATGAAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db 241 CACATGAAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
QY 301 GAACGCAATGATGAAGTAATTTAATGTTCAACTAACCTTTAGGAGAAAATATGAACA 360
Db 301 GAACGCAATGATGAAGTAATTTAATGTTCAACTAACCTTTAGGAGAAAATATGAACA 360
QY 361 ATATATCGTCTCAAAATTCAGCAAAACGCTGTAATCTTTGGTGTCTGTCTGTAATGGCA 420
Db 361 ATATATCGTCTCAAAATTCAGCAAAACGCTGTAATCTTTGGTGTCTGTCTGTAATGGCA 420
QY 421 CGGGGTGTGACCATTCACAGAAAAGGCTTCGGCTATGTTACTATCTTTTAGGTGTAAC 480
Db 421 CGGGGTGTGACCATTCACAGAAAAGGCTTCGGCTATGTTACTATCTTTTAGGTGTAAC 480
QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTC 540
Db 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTC 540
QY 541 CAATCTGTTTACGACGCGCTTACAAGAAATGATGATGATGATGATGATGATGATGATG 600
Db 541 CAATCTGTTTACGACGCGCTTACAAGAAATGATGATGATGATGATGATGATGATGATG 600
QY 601 CAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
Db 601 CAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
QY 661 CAATTTAATATCGACCAAAATGAAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
Db 661 CAATTTAATATCGACCAAAATGAAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
QY 721 GTATTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGATTTTAGATTTTAC 780
Db 721 GTATTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGATTTTAGATTTTAC 780
QY 781 GGCAAGTCTTTTAAATCAACCCCAATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 781 GGCAAGTCTTTTAAATCAACCCCAATGATGATGATGATGATGATGATGATGATGATGAT 840

```



QY 841 ACTAATGGCTTTACGGCTTCTACGGTAGACATTTCTACGAAACATCAAGCGCGTAAT 900  
DB 841 ACTAATGGCTTTACGGCTTCTACGGTAGACATTTCTACGAAACATCAAGCGCGTAAT 900  
QY 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACGGTTTAAT 960  
DB 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATCACGGTTTAAT 960  
QY 961 ACTGTCGGTAAAGACGGCAGTGTAAATCTTATTTGGTGGCAAGTGAAGAACAGGGGTG 1020  
DB 961 ACTGTCGGTAAAGACGGCAGTGTAAATCTTATTTGGTGGCAAGTGAAGAACAGGGGTG 1020  
QY 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTGCGAGGCAAAATACACCATCAGCGAT 1080  
DB 1021 ATTAGCGTAAATGGTGGCAGCATTTCTTACTGCGAGGCAAAATACACCATCAGCGAT 1080  
QY 1081 ATAATAACCCCAACCATTTACTTACAGCATTTGCGCGCCCTGAAATGAAGCGGTCAATCTG 1140  
DB 1081 ATAATAACCCCAACCATTTACTTACAGCATTTGCGCGCCCTGAAATGAAGCGGTCAATCTG 1140  
QY 1141 GCGGATATTTTGGCAAGCGGTAAACATTAATGTCGCTGCCACTATTCGAACCA 1200  
DB 1141 GCGGATATTTTGGCAAGCGGTAAACATTAATGTCGCTGCCACTATTCGAACCA 1200  
QY 1201 GGTAACTTTCTGCTGATCTGTAAGCAAGATAAAGCGGCAATATTTGTTCTCCGC 1260  
DB 1201 GGTAACTTTCTGCTGATCTGTAAGCAAGATAAAGCGGCAATATTTGTTCTCCGC 1260  
QY 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320  
DB 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320  
QY 1321 GCGAAGCTGATGATTTACAGGCGATAAAGTCAATTTAAAAACAGGTGCGATTCGACCTT 1380  
DB 1321 GCGAAGCTGATGATTTACAGGCGATAAAGTCAATTTAAAAACAGGTGCGATTCGACCTT 1380  
QY 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTTGGCGGTGACGAGCGCGCGAAGGTAAAAAC 1440  
DB 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTTGGCGGTGACGAGCGCGCGAAGGTAAAAAC 1440  
QY 1441 GGCATCAATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATGATATCAGC 1500  
DB 1441 GGCATCAATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATGATATCAGC 1500  
QY 1501 AAAGAAAGCGGACGCGCTATTGTGGGCGCATATTGGCTTAATTTGACGGCAATTT 1560  
DB 1501 AAAGAAAGCGGACGCGCTATTGTGGGCGCATATTGGCTTAATTTGACGGCAATTT 1560  
QY 1561 AACGCTCAAGGTAGTGGTATCGCTAAACCGGTGTTTGTGGAGACATCGGGCAT 1620  
DB 1561 AACGCTCAAGGTAGTGGTATCGCTAAACCGGTGTTTGTGGAGACATCGGGCAT 1620  
QY 1621 TATTTATCCATTTGACGCAATGCAATTTTAAACAAAGAGTGGTGTAGACCCCTGAT 1680  
DB 1621 TATTTATCCATTTGACGCAATGCAATTTTAAACAAAGAGTGGTGTAGACCCCTGAT 1680  
QY 1681 GATGTAACAATTTGAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGAATTC 1740  
DB 1681 GATGTAACAATTTGAAGCCGAAGACCCCTTCGCAATAATACCGGTATAAATGATGAATTC 1740  
QY 1741 CCAACAGGCGGTTGAAGCAAGCGACCTTAAAAAATAGCGAACTCAAAACACGCTA 1800  
DB 1741 CCAACAGGCGGTTGAAGCAAGCGACCTTAAAAAATAGCGAACTCAAAACACGCTA 1800  
QY 1801 ACCAATACACATTTTCAAAATTTCTGAAAACGCTGACAAATGAATTAACGGCATCA 1860  
DB 1801 ACCAATACACATTTTCAAAATTTCTGAAAACGCTGACAAATGAATTAACGGCATCA 1860  
QY 1861 AGAAAACCTTACGGTAAATAGCTCAATCAACATCGGACCACTCCCACTTAATTCCTCAT 1920  
DB 1861 AGAAAACCTTACGGTAAATAGCTCAATCAACATCGGACCACTCCCACTTAATTCCTCAT 1920  
QY 1921 ACTAAAGGTCAGCGTGGCGGAGGCGTTTCAGATTGATGGAGATATTACTTCTAAAGCGGA 1980

DB 1921 AGTAAAGTCAGCCTGGCGGAGGCGTTTCAGATTGATGAGATATTACTTCTAAAGCGGA 1980  
QY 1981 AATTTAACCATTTATCTGGCGGATGGTTCATCTTCAAAAAATATTAGCTTTGATCAG 2040  
DB 1981 AATTTAACCATTTATCTGGCGGATGGTTCATCTTCAAAAAATATTAGCTTTGATCAG 2040  
QY 2041 GGTTTTTTAAATATACCGCGCTTCCTAGCTTTTGAAGTGGAAATTAACAAAGCACGC 2100  
DB 2041 GGTTTTTTAAATATACCGCGCTTCCTAGCTTTTGAAGTGGAAATTAACAAAGCACGC 2100  
QY 2101 GACGGGCAATATGCTAAATTTGTCGCCAGGCGACTGTAAACATTTACAGGAGGGA 2160  
DB 2101 GACGGGCAATATGCTAAATTTGTCGCCAGGCGACTGTAAACATTTACAGGAGGGA 2160  
QY 2161 GATTTACGGGTAAACACGATATCTTTAAACGGGTAAGGTCTGAATATCATTTCA 2220  
DB 2161 GATTTACGGGTAAACACGATATCTTTAAACGGGTAAGGTCTGAATATCATTTCA 2220  
QY 2221 TCAGTGAATTAATTTAACCCACAATCTTAGTGCAATTAACATATCTGGGAATAACA 2280  
DB 2221 TCAGTGAATTAATTTAACCCACAATCTTAGTGCAATTAACATATCTGGGAATAACA 2280  
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGCATGG 2340  
DB 2281 ATTAACCAAACTACGAGAAAGAACACCTCGTATTGGCAAAACAGCCATGATTCGCATGG 2340  
QY 2341 AACGTCAGTCTTAACTAGAGACAGCGCAAAATTTTACCTTTATTAATACATTTCA 2400  
DB 2341 AACGTCAGTCTTAACTAGAGACAGCGCAAAATTTTACCTTTATTAATACATTTCA 2400  
QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAAGTCTGACAGGCTGAATTTTAAACGC 2460  
DB 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAAGTCTGACAGGCTGAATTTTAAACGC 2460  
QY 2461 GTAAATGGCAACATGTCATTCATCTCAAGAGGAGGCAAAAGTTAAATTTCAATTTAAA 2520  
DB 2461 GTAAATGGCAACATGTCATTCATCTCAAGAGGAGGCAAAAGTTAAATTTCAATTTAAA 2520  
QY 2521 CCAACGAGCAATGATCAACCAAGCAACCTTTTACCAATTCGGTTTACGCAATATCA 2580  
DB 2521 CCAACGAGCAATGATCAACCAAGCAACCTTTTACCAATTCGGTTTACGCAATATCA 2580  
QY 2581 GCCACTGGTGGGCGCTCTGTTTTTGTATATATATGCAACCATTTCTGGCAGAGGGCT 2640  
DB 2581 GCCACTGGTGGGCGCTCTGTTTTTGTATATATATGCAACCATTTCTGGCAGAGGGCT 2640  
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTACGGCGCTAAATTTTACCTTTAAATTTCCCAT 2700  
DB 2641 GAGTTAAAAATGAGTGAATTAATATCTCTACGGCGCTAAATTTTACCTTTAAATTTCCCAT 2700  
QY 2701 GTTCGGCGGATGACGCTTTTAAAAATCAACAAAGACTTAAACCAATAATGCAACCAATTC 2760  
DB 2701 GTTCGGCGGATGACGCTTTTAAAAATCAACAAAGACTTAAACCAATAATGCAACCAATTC 2760  
QY 2761 AATTTACGCTCAGACAGCAAGAGTATTTTATGACGGGTACGACCAATGCGCATC 2820  
DB 2761 AATTTACGCTCAGACAGCAAGAGTATTTTATGACGGGTACGACCAATGCGCATC 2820  
QY 2821 AATTTACGCTCAGACATATCCATTTCTGGGCGTAAATGTCACCTTTGGTGAACAACTCA 2880  
DB 2821 AATTTACGCTCAGACATATCCATTTCTGGGCGTAAATGTCACCTTTGGTGAACAACTCA 2880  
QY 2881 AGCAGCAGCATTCAGGGGAATATTTATCTAGAAAGCAGCAATTTAGCTAGAGCC 2940  
DB 2881 AGCAGCAGCATTCAGGGGAATATTTATCTAGAAAGCAGCAATTTAGCTAGAGCC 2940  
QY 2941 AATAACGCCCTTAATCAGCAAAACATAAGGATAGAGTTTAAACCTTTGGCAGCTTCCCTC 3000  
DB 2941 AATAACGCCCTTAATCAGCAAAACATAAGGATAGAGTTTAAACCTTTGGCAGCTTCCCTC 3000  
QY 3001 GTTAATGGAGCTTTAAGTTTAACTGGGGAATTCAGATTTAAAGGCAATCTCCTATT 3060  
DB 3001 GTTAATGGAGCTTTAAGTTTAACTGGGGAATTCAGATTTAAAGGCAATCTCCTATT 3060





Db 1681 GATGTACAAATTTGAAGCGAGAGACCCCTTCGCAATTAATACCGGTATAAATGAATTC 1740  
QY 1741 CCAACAGCCACCGGTGAGCAAGGAGCCCTAAAAAATAGCGAATCAAAACACGCTA 1800  
Db 1741 CCAACAGCCACCGGTGAGCAAGGAGCCCTAAAAAATAGCGAATCAAAACACGCTA 1800  
QY 1801 ACCAATAAATTTTCAAAATATCTGAAAGAGCGCTGAGCAATGAATTAACGGCATCA 1860  
Db 1801 ACCAATAAATTTTCAAAATATCTGAAAGAGCGCTGAGCAATGAATTAACGGCATCA 1860  
QY 1861 AGAAATCTTACCGTTAATAGCTCAATCAACATCGGAAGCAATCCCACTTAATTCGCAT 1920  
Db 1861 AGAAATCTTACCGTTAATAGCTCAATCAACATCGGAAGCAATCCCACTTAATTCGCAT 1920  
QY 1921 AGTAAGCTCAGCGTGGGAGGCGTTTCAGATGATGGAGATATTAATCTTAAAGGGGA 1980  
Db 1921 AGTAAGCTCAGCGTGGGAGGCGTTTCAGATGATGGAGATATTAATCTTAAAGGGGA 1980  
QY 1981 AATTTAACCAATTTATCTGGCGGATGGGTGATGTTTCATATAAATATTAACGTTGATCAG 2040  
Db 1981 AATTTAACCAATTTATCTGGCGGATGGGTGATGTTTCATATAAATATTAACGTTGATCAG 2040  
QY 2041 GGTGTTTTTAATATTTACCGCGCTTCGCTAGCTTTTGAAGGTGGAATTAACAAAGCAGC 2100  
Db 2041 GGTGTTTTTAATATTTACCGCGCTTCGCTAGCTTTTGAAGGTGGAATTAACAAAGCAGC 2100  
QY 2101 GACGCGCAATATGCTAAAAATTTGCGCCAGGCGACTGTAAACCATTAACAGGAGGAAAA 2160  
Db 2101 GACGCGCAATATGCTAAAAATTTGCGCCAGGCGACTGTAAACCATTAACAGGAGGAAAA 2160  
QY 2161 GATTTACAGGCTAACCAAGTATCTTTAAACGCAAGCGGTAAAGGCTGGAATATCAATTTCA 2220  
Db 2161 GATTTACAGGCTAACCAAGTATCTTTAAACGCAAGCGGTAAAGGCTGGAATATCAATTTCA 2220  
QY 2221 TCAGTGAATTAATTTAACCCCAATCTTAGTGCCACAATTAACATATCTGGGAATATAACA 2280  
Db 2221 TCAGTGAATTAATTTAACCCCAATCTTAGTGCCACAATTAACATATCTGGGAATATAACA 2280  
QY 2281 ATTAACCAATACGAGAGAAAGACCTGATATGCGCAACACCAAGCCATGATTCGCACTGG 2340  
Db 2281 ATTAACCAATACGAGAGAAAGACCTGATATGCGCAACACCAAGCCATGATTCGCACTGG 2340  
QY 2341 AAGTCAGTGTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2400  
Db 2341 AAGTCAGTGTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2400  
QY 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGCTCTGAGGGGTGAATTTAACGGC 2460  
Db 2401 AGCAATAGCAAGGCTTAACACACAGTATAGAGCTCTGAGGGGTGAATTTAACGGC 2460  
QY 2461 GTAATGCAACATGTCATTAATCTCAAGAGGAGCGGAAGTTAATTTCAATTTAAAA 2520  
Db 2461 GTAATGCAACATGTCATTAATCTCAAGAGGAGCGGAAGTTAATTTCAATTTAAAA 2520  
QY 2521 CCAAGGAGCAATGAACACAGCAACACCTTTACCAATTCGGTTTTTACCAATATATCACA 2580  
Db 2521 CCAAGGAGCAATGAACACAGCAACACCTTTACCAATTCGGTTTTTACCAATATATCACA 2580  
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTATATATATGCAACCATTTCTGCGAGGGGCT 2640  
Db 2581 GCCACTGGTGGGGCTCTGTTTTTTTATATATATGCAACCATTTCTGCGAGGGGCT 2640  
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGCGCTTAATTTTACCTTTAAATTCCTAT 2700  
Db 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGCGCTTAATTTTACCTTTAAATTCCTAT 2700  
QY 2701 GTTCCGGCGATGACGCTTTTAAATACAAAGAGCTTAACCAATTAATGCAACCAATTTCA 2760  
Db 2701 GTTCCGGCGATGACGCTTTTAAATACAAAGAGCTTAACCAATTAATGCAACCAATTTCA 2760  
QY 2761 AATTTACGCTCAGACAGCAAGAGATGATTTTTTATGCGGTACGCAAGCAATGCAATC 2820  
Db 2761 AATTTACGCTCAGACAGCAAGAGATGATTTTTTATGCGGTACGCAAGCAATGCAATC 2820

QY 2821 AATTTACACCTACACATATTCATTTCTGGCGGTAAATGTACCCCTTGGTGAGAAAACTCA 2880  
Db 2821 AATTTACACCTACACATATTCATTTCTGGCGGTAAATGTACCCCTTGGTGAGAAAACTCA 2880  
QY 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAAGCAGCAATGTACGCTAGAGCC 2940  
Db 2881 AGCAGCAGCATTTACGGGAATATTACTATCGAGAAAGCAGCAATGTACGCTAGAGCC 2940  
QY 2941 AATAAGCCCTTAATCAGCAAAACATAGGATAGAGTTAATAAATCTGGAGCTTGGCTC 3000  
Db 2941 AATAAGCCCTTAATCAGCAAAACATAGGATAGAGTTAATAAATCTGGAGCTTGGCTC 3000  
QY 3001 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCAATATTAAGGCAATCTCACTATT 3060  
Db 3001 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCAATATTAAGGCAATCTCACTATT 3060  
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCCCTAAATATACCGGCAATTTT 3120  
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAGACTAGAGATACCCCTAAATATACCGGCAATTTT 3120  
QY 3121 ACCAATAATGGCAGCTGCGGAATTAATATAACAAAGAGTGGTAAATCTTGGCAATGTT 3180  
Db 3121 ACCAATAATGGCAGCTGCGGAATTAATATAACAAAGAGTGGTAAATCTTGGCAATGTT 3180  
QY 3181 ACCAATGATGTTGATTTAAACATTAACACTACAGCTAAACGCAAGAGTGGTAAATCTTGGCAATGTT 3180  
Db 3181 ACCAATGATGTTGATTTAAACATTAACACTACAGCTAAACGCAAGAGTGGTAAATCTTGGCAATGTT 3180  
QY 3241 GCGGAGATATATCAACAAAAAGGAGCTTAAATATTAACAAAGAGTGGTAAATCTTGGCAATGTT 3300  
Db 3241 GCGGAGATATATCAACAAAAAGGAGCTTAAATATTAACAAAGAGTGGTAAATCTTGGCAATGTT 3300  
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTTCGAT 3360  
Db 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTTCGAT 3360  
QY 3361 AANAATATATCAACAAAGAGTAACTAAACAAAGAGTAACTAAACAAAGAGTAACTAACTAGT 3420  
Db 3361 AANAATATATCAACAAAGAGTAACTAAACAAAGAGTAACTAAACAAAGAGTAACTAACTAGT 3420  
QY 3421 TCAGATGCGCAAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3480  
Db 3421 TCAGATGCGCAAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3480  
QY 3481 GACCTAAGTATTTTCAAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3540  
Db 3481 GACCTAAGTATTTTCAAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3540  
QY 3541 ACTATTGCAACAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3600  
Db 3541 ACTATTGCAACAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3600  
QY 3601 AATGTTAAAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3660  
Db 3601 AATGTTAAAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3660  
QY 3661 AANAATCTAGCAGCAATGGCGAGTAAAGCAATAGCAGCAATAGCAGCAATAGCAGCAATAGCAGCA 3720  
Db 3661 AANAATCTAGCAGCAATGGCGAGTAAAGCAATAGCAGCAATAGCAGCAATAGCAGCAATAGCAGCA 3720  
QY 3721 ATTACTGCAAAAAATAGAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3780  
Db 3721 ATTACTGCAAAAAATAGAGTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT 3780  
QY 3781 ACCGCTCGGAAAGGTTACCACAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3840  
Db 3781 ACCGCTCGGAAAGGTTACCACAGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 3840  
QY 3841 GCAAGTATTAACAAACAAAGGTTATACGCGGTACGATTTCCGGTAAACAGGTAAGT 3900  
Db 3841 GCAAGTATTAACAAACAAAGGTTATACGCGGTACGATTTCCGGTAAACAGGTAAGT 3900

3901	Qy	GTTAGCGCGACTGGTGTATTTAAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATCGGGT	3960
3901	Db	GTTAGCGCGACTGGTGTATTTAAACCACTAAATCCGGCTCAAAATTTGAAGCGAAATCGGGT	3960
3961	Qy	GAGGCTAATGTAAACAAGTGCACAAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATACGGTA	4020
3961	Db	GAGGCTAATGTAAACAAGTGCACAAGGTACAAATTTGGCGGTACAAATTTCCGGTAAATACGGTA	4020
4021	Qy	AATGTTACGGCAAAACGCTGGCGATTTTAAACAGTTTGGGAATGGCGACGAAATTAATCGGACA	4080
4021	Db	AATGTTACGGCAAAACGCTGGCGATTTTAAACAGTTTGGGAATGGCGACGAAATTAATCGGACA	4080
4081	Qy	GAAGGAGCTGCAACCTTTAAACCGCAACAGGAAATACCTTGACTACTGAAGCCGGTCTAGC	4140
4081	Db	GAAGGAGCTGCAACCTTTAAACCGCAACAGGAAATACCTTGACTACTGAAGCCGGTCTAGC	4140
4141	Qy	ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCGCAGGAAGC	4200
4141	Db	ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGAAATGCTAGCATCGCAGGAAGC	4200
4201	Qy	ATTAAATGCTGATATGTGACATTTAAATACATACAGGACACTTAACACCGTGGCAGGCTCG	4260
4201	Db	ATTAAATGCTGATATGTGACATTTAAATACATACAGGACACTTAACACCGTGGCAGGCTCG	4260
4261	Qy	GATATTTAAAGCAACAGCGGCACCTTGGTTATTTAAACGCAAAAGATGCTAAGCTAAATGGT	4320
4261	Db	GATATTTAAAGCAACAGCGGCACCTTGGTTATTTAAACGCAAAAGATGCTAAGCTAAATGGT	4320
4321	Qy	GATGCATCAGGTGATGATGACAGAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG	4380
4321	Db	GATGCATCAGGTGATGATGACAGAAGTGAATGCAGTCAACGCAAGCGGCTCTGGTAGTGTG	4380
4381	Qy	ACTGCGGCAACCTCAAGCAGTGTGGAATATCACTCGGGGATTTAAACACAGTAAATGGTTA	4440
4381	Db	ACTGCGGCAACCTCAAGCAGTGTGGAATATCACTCGGGGATTTAAACACAGTAAATGGTTA	4440
4441	Qy	AATATCATTTCCGAAAGATGGTAGAAACACTCTGCGCTTAAAGAGCAAGGAAATTGAGGTG	4500
4441	Db	AATATCATTTCCGAAAGATGGTAGAAACACTCTGCGCTTAAAGAGCAAGGAAATTGAGGTG	4500
4501	Qy	AAATATATCCAGCCAGGTGTAGCAAGTGTAGAAGAAGTAAATTGAAGCGAAACCGCTCCTT	4560
4501	Db	AAATATATCCAGCCAGGTGTAGCAAGTGTAGAAGAAGTAAATTGAAGCGAAACCGCTCCTT	4560
4561	Qy	GAAAAAGTAAAGATTTATCTGATCAAGCAAAAGAGAAACATTTAGCTAAACTTGGTGTAAAGT	4620
4561	Db	GAAAAAGTAAAGATTTATCTGATCAAGCAAAAGAGAAACATTTAGCTAAACTTGGTGTAAAGT	4620
4621	Qy	GCTGTACGTTTTGTTGAGCCAAATTAATCAATTCAGTCAATACACAAAATGAATTTTACA	4680
4621	Db	GCTGTACGTTTTGTTGAGCCAAATTAATCAATTCAGTCAATACACAAAATGAATTTTACA	4680
4681	Qy	ACCAGACCGTCAAGTCAAGTGATATTTCTGAAGTAAGGCGTGTCTTCTCAAGTGGTAAAT	4740
4681	Db	ACCAGACCGTCAAGTCAAGTGATATTTCTGAAGTAAGGCGTGTCTTCTCAAGTGGTAAAT	4740
4741	Qy	GGCGCAGGATGTATACCAATCTTCTGCTGACGATGAGACGCGTAGTCAGTAAATTCACAAG	4800
4741	Db	GGCGCAGGATGTATACCAATCTTCTGCTGACGATGAGACGCGTAGTCAGTAAATTCACAAG	4800
4801	Qy	GTAGATTTCACTCGAATGAAGTCATTTATTTTCGTATTATTTACTGTGTGGGTTAAA	4860
4801	Db	GTAGATTTCACTCGAATGAAGTCATTTATTTTCGTATTATTTACTGTGTGGGTTAAA	4860
4861	Qy	GTTACAGTACGGGCTTTACCCATCTTGTAAAAAATTTACGGAGATACAAATAAAGTATTTTT	4920
4861	Db	GTTACAGTACGGGCTTTACCCATCTTGTAAAAAATTTACGGAGATACAAATAAAGTATTTTT	4920
4921	Qy	AACAGGTTATTATTATG	4937
4921	Db	AACAGGTTATTATTATG	4937

```

1 RESULT 4
2 US 08-469-880-3
3 Sequence 3, Application US/08469880
4 Patent No. 5876733
5 GENERAL INFORMATION:
6 APPLICANT: Barokamp, Stephen J.
7 TITLE OF INVENTION: High Molecular Weight Surface Proteins
8 TITLE OF INVENTION: Of No. 5876733-Typeable Haemophilus
9 NUMBER OF SEQUENCES: 8
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Shoemaker and Mattare, Ltd.
12 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
13 STREET: Bldg. 1
14 CITY: Arlington
15 STATE: Virginia
16 COUNTRY: U.S.A.
17 ZIP: 22202-0286
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC Compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/469-880
25 FILING DATE: 06-JUN-1995
26 CLASSIFICATION: 435
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: GB 9205704.1
29 FILING DATE: 16-MAR-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US PCT/US93/02166
32 FILING DATE: 16-MAR-1993
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/302,832
35 FILING DATE: 16-SEP-1994
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Berkstresser, Jerry W
38 REGISTRATION NUMBER: 22,651
39 REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: (703) 415-0810
42 TELEFAX: (703) 415-0813
43 INFORMATION FOR SEQ ID NO: 3:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 4937 base pairs
46 TYPE: nucleic acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49 MOLECULE TYPE: DNA (genomic)

```

Query Match	100.0%	Score 4937;	DB 2;	Length 4937;
Best Local Similarity	100.0%;	pred. NO. 0;		
Matches 4937; Conservative	0%;	Mismatches	0;	Gaps 0;
		Indels	0;	





Qy	2461	GTAATPGCAACATGTCAATTC	AAATCTCAAGAGGAGCGAAAGTTAA	TTTCAAAATTA	AA	2520	
Db	2461	GTAATPGCAACATGTCAATTC	AAATCTCAAGAGGAGCGAAAGTTAA	TTTCAAAATTA	AA	2520	
Qy	2521	CCAAAGGAACATGAACACAA	AGCAACCTTTACCAATCGGTTTT	TAGCAATATCA	CA	2580	
Db	2521	CCAAAGGAACATGAACACAA	AGCAACCTTTACCAATCGGTTTT	TAGCAATATCA	CA	2580	
Qy	2581	GCACGTGGGGGCTGCTTTTT	TTTGGATATATGCCAACCTCT	GGCAGAGGGCT		2640	
Db	2581	GCACGTGGGGGCTGCTTTTT	TTTGGATATATGCCAACCTCT	GGCAGAGGGCT		2640	
Qy	2641	GAGTTAAAAATGAGTGA	AAATTAATCTCTAACGCGCTAA	TTTACCTTTAAAT	CCCCAT	2700	
Db	2641	GAGTTAAAAATGAGTGA	AAATTAATCTCTAACGCGCTAA	TTTACCTTTAAAT	CCCCAT	2700	
Qy	2701	GTTTCGGCGGATGACGCT	TTTTAAAAATCAACAAGACT	TTAACCATAAATGCA	CAACCAATTC	2760	
Db	2701	GTTTCGGCGGATGACGCT	TTTTAAAAATCAACAAGACT	TTAACCATAAATGCA	CAACCAATTC	2760	
Qy	2761	AATTTTCAGCCTCAGAC	AGCAAGACATGATTTTATGACGGT	ACGACACCAATGC	CAATC	2820	
Db	2761	AATTTTCAGCCTCAGAC	AGCAAGACATGATTTTATGACGGT	ACGACACCAATGC	CAATC	2820	
Qy	2821	AATTCACCTCAACATAT	CCACTCTGGCGGTAAATG	CTACCCCTGGTGAC	CAAAAATTC	2880	
Db	2821	AATTCACCTCAACATAT	CCACTCTGGCGGTAAATG	CTACCCCTGGTGAC	CAAAAATTC	2880	
Qy	2881	AGCAGACGATTTACGGG	GAATTAATCTCGAGAAAGCAG	CAAAATGTTACGTTAG	AGCC	2940	
Db	2881	AGCAGACGATTTACGGG	GAATTAATCTCGAGAAAGCAG	CAAAATGTTACGTTAG	AGCC	2940	
Qy	2941	AATAAGCCCCTAATC	AGCAAAACATAAGGATAG	ATTATAAACCTTGGC	AGCTTGCTC	3000	
Db	2941	AATAAGCCCCTAATC	AGCAAAACATAAGGATAG	ATTATAAACCTTGGC	AGCTTGCTC	3000	
Qy	3001	GTTAATGGGATTTAAG	TTTAATCTGGCGAAATCG	AGATATTAAGGCAAT	CTCACTATT	3060	
Db	3001	GTTAATGGGATTTAAG	TTTAATCTGGCGAAATCG	AGATATTAAGGCAAT	CTCACTATT	3060	
Qy	3061	TCAGAAAGCGCCACT	TTTAAAGGAAAGACTAG	AGATACCTTAAATAT	CACCGGCAATTTT	3120	
Db	3061	TCAGAAAGCGCCACT	TTTAAAGGAAAGACTAG	AGATACCTTAAATAT	CACCGGCAATTTT	3120	
Qy	3121	ACCAATATGGCACTG	CCGAAAATTAATTAAC	ACAAGAGTGGTAA	AACTTGGCAATGTT	3180	
Db	3121	ACCAATATGGCACTG	CCGAAAATTAATTAAC	ACAAGAGTGGTAA	AACTTGGCAATGTT	3180	
Qy	3181	ACCAATGATGGTATTA	AAACATTAACCTACG	CTAAACGCAACCA	AAAGCATCATC	3240	
Db	3181	ACCAATGATGGTATTA	AAACATTAACCTACG	CTAAACGCAACCA	AAAGCATCATC	3240	
Qy	3241	GGCGGAGATATATCA	CAAAAAGAGCTTAA	ATATACAGACAGT	TAATTAATGATGCT	3300	
Db	3241	GGCGGAGATATATCA	CAAAAAGAGCTTAA	ATATACAGACAGT	TAATTAATGATGCT	3300	
Qy	3301	GAATCCAAATTTGGG	CGCAATCTCGCAAAAG	AGCGCAACCTTC	TCATCCGAT	3360	
Db	3301	GAATCCAAATTTGGG	CGCAATCTCGCAAAAG	AGCGCAACCTTC	TCATCCGAT	3360	
Qy	3361	AAATTAATATCA	CAAAACAGATACA	ATCAAAAAGGTTAT	GATGAGAGGACTCTAGT	3420	
Db	3361	AAATTAATATCA	CAAAACAGATACA	ATCAAAAAGGTTAT	GATGAGAGGACTCTAGT	3420	
Qy	3421	TCAGATGCGCAAGT	TAATGCCAACCTTA	CTATTAAAA	CAAGAATTTGAAAT	TCGACAGAA	3480
Db	3421	TCAGATGCGCAAGT	TAATGCCAACCTTA	CTATTAAAA	CAAGAATTTGAAAT	TCGACAGAA	3480
Qy	3481	GACCTAAGTATTT	CAGTTTTCAATAAG	CAGAGATTACAGCA	AGATGGTAGAGTTTA	3540	
Db	3481	GACCTAAGTATTT	CAGTTTTCAATAAG	CAGAGATTACAGCA	AGATGGTAGAGTTTA	3540	
Qy	3541	ACTATTGGCAACAG	TAAATGACGCTAC	ACGCGTCCGGAAG	CAAAACAGTAAC	TTTAAAC	3600

[illegible]

Db 4621 GCTGTACGTTTGTGAGCCAAATATACAAATACAGTCAATACACAAATGAATTTACA 4680  
Qy 4681 ACCGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAT 4740  
Db 4681 ACCGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTGTTCTCAAGTGGTAAT 4740  
Qy 4741 GCGGACGAGTATGACCAATGTTGCTGACGATGGACGCCGTAGTCAATGACAAAG 4800  
Db 4741 GCGGACGAGTATGACCAATGTTGCTGACGATGGACGCCGTAGTCAATGACAAAG 4800  
Qy 4801 GAGATTTTCATCCTCAATGAAGTCAATTTATTTTCGTAATTTACTGTTGGGTAAA 4860  
Db 4801 GAGATTTTCATCCTCAATGAAGTCAATTTATTTTCGTAATTTACTGTTGGGTAAA 4860  
Qy 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAATAAAGTATTTT 4920  
Db 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAATAAATACGAGAAATACAATAAAGTATTTT 4920  
Qy 4921 AACAGGTTATTATTATG 4937  
Db 4921 AACAGGTTATTATTATG 4937

## RESULT 5

US-08-728-470-3  
; Sequence 3, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstreser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4937 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-728-470-3

Query Match 100.0%; Score 4937; DB 2; Length 4937;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TAAATATACAAAGATATAAATAAATCAAGATTTTGTGATGACAAACAACAATTTACAA 60  
Db 1 TAAATATACAAAGATATAAATAAATCAAGATTTTGTGATGACAAACAACAATTTACAA 60  
Qy 61 CACCTTTTGTGAGTCTATATGCAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120  
Db 61 CACCTTTTGTGAGTCTATATGCAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120  
Qy 121 AATGTAATATCTTTTCATCTTTTCACTTTTAACTTTTCATCTTTTCATCTTTTCATCTT 180  
Db 121 AATGTAATATCTTTTCATCTTTTCACTTTTAACTTTTTCATCTTTTCATCTTTTCATCTT 180  
Qy 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
Db 181 CTTTCATCTTTTCATCTTTTCACTTTTCACTTTTTCATCTTTTCATCTTTTCATCTTT 240  
Qy 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Qy 301 GAAACGCAAAATGATAAAGTAAATTTTAACTTAACTTAACTTAACTTAACTTAACTTAA 360  
Db 301 GAAACGCAAAATGATAAAGTAAATTTTAACTTAACTTAACTTAACTTAACTTAACTTAA 360  
Qy 361 ATATATCTGTCACAAATTCAGCAAAAGCGCTTAACTTAACTTAACTTAACTTAACTTAA 420  
Db 361 ATATATCTGTCACAAATTCAGCAAAAGCGCTTAACTTAACTTAACTTAACTTAACTTAA 420  
Qy 421 CGGGTCTGACCAATTCACAGAAAGCGCTTCCGCTATGTTAGTGTACATCTTTTAGGTGTA 480  
Db 421 CGGGTCTGACCAATTCACAGAAAGCGCTTCCGCTATGTTAGTGTACATCTTTTAGGTGTA 480  
Qy 481 CACITAGCGTTAAAGCGCTTCCGCTATGTTAGTGTACATCTTTTAGGTGTACATCTTTAG 540  
Db 481 CACITAGCGTTAAAGCGCTTCCGCTATGTTAGTGTACATCTTTTAGGTGTACATCTTTAG 540  
Qy 541 CAATCTGTTTAGCAAGCGGCTTCAAGGAATGGATGTAGTACACGGCACAGCCACTATG 600  
Db 541 CAATCTGTTTAGCAAGCGGCTTCAAGGAATGGATGTAGTACACGGCACAGCCACTATG 600  
Qy 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTTAATTTGAAA 660  
Db 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTTAATTTGAAA 660  
Qy 661 CAATTTAATACATCGACCAAAATGAAATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 720  
Db 661 CAATTTAATACATCGACCAAAATGAAATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 720  
Qy 721 GTATTCAACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTAGATTTTAAAC 780  
Db 721 GTATTCAACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTAGATTTTAAAC 780  
Qy 781 GGACAAGTCTTTTAAATCAACCGGCTTCAAGGAATGGATGTAGTACACGGCACAGCCACT 840  
Db 781 GGACAAGTCTTTTAAATCAACCGGCTTCAAGGAATGGATGTAGTACACGGCACAGCCACT 840  
Qy 841 ACTAATGCGTTTACGGCTTCTACGCTAGACATTTCTAACGAAAAATCAAGCGCGGTAA 900  
Db 841 ACTAATGCGTTTACGGCTTCTACGCTAGACATTTCTAACGAAAAATCAAGCGCGGTAA 900  
Qy 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATTCACGGTTTAA 960  
Db 901 TTCACCTTCGAGCAAAACCAAGATAAAGCGCTCGCTGAAATTTGTGAATTCACGGTTTAA 960  
Qy 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTTGTCGCAAGTGAAGGAGGAGGAGG 1020  
Db 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTTGTCGCAAGTGAAGGAGGAGGAGG 1020



Db 3181 ACCAATGATGGTGAATTAACCAATTACCACTACCGCTAAACGCAACCAAGAACGATCATC 3240  
QY 3241 GCGGAGATATATCAACAAAGAAAGCTTAATAATATACAGACAGTAATAATGATGCT 3300  
Db 3241 GCGGAGATATATCAACAAAGAAAGCTTAATAATATACAGACAGTAATAATGATGCT 3300  
QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAAAGCAACCTTCCTCCGAT 3360  
Db 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAGAAAGCAACCTTCCTCCGAT 3360  
QY 3361 AAAATTAATATACCAACACAGATACAAATCAAAAGGATTTGATGAGAGGACTCTAGT 3420  
Db 3361 AAAATTAATATACCAACACAGATACAAATCAAAAGGATTTGATGAGAGGACTCTAGT 3420  
QY 3421 TCAGATGCGACAACTAATGCGCACTAATCTTAACCAAGAAATGAAATGACAGAA 3480  
Db 3421 TCAGATGCGACAACTAATGCGCACTAATCTTAACCAAGAAATGAAATGACAGAA 3480  
QY 3481 GACCTAAGTATTTAGGTTTCAATAAGCAGAGATTACAGCCAAAGATGGTAGAGATTGA 3540  
Db 3481 GACCTAAGTATTTAGGTTTCAATAAGCAGAGATTACAGCCAAAGATGGTAGAGATTGA 3540  
QY 3541 ACTATTGGCAACAGTAAAGCGGTAAACAGCGTGCAGGCAAGCAACAGTAACCTTTAAC 3600  
Db 3541 ACTATTGGCAACAGTAAAGCGGTAAACAGCGTGCAGGCAAGCAACAGTAACCTTTAAC 3600  
QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAACAATGTGACACTAAATAGCAAGTG 3660  
Db 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAACAATGTGACACTAAATAGCAAGTG 3660  
QY 3661 AAAACATCTAGCAGCAATGCGGACCTGAAAGCAATAGCAGCAACAGTACCGGCTTAATC 3720  
Db 3661 AAAACATCTAGCAGCAATGCGGACCTGAAAGCAATAGCAGCAACAGTACCGGCTTAATC 3720  
QY 3721 ATTACTGCAAAAATTTAGAGTAAACAAAGATATTACTTCTCAAAACAGTAATAATATC 3780  
Db 3721 ATTACTGCAAAAATTTAGAGTAAACAAAGATATTACTTCTCAAAACAGTAATAATATC 3780  
QY 3781 ACCGCTCGGAAAGGTTTACCAACACAGCAGCGCTCGACCATTAACGCAACAAATGGCAAA 3840  
Db 3781 ACCGCTCGGAAAGGTTTACCAACACAGCAGCGCTCGACCATTAACGCAACAAATGGCAAA 3840  
QY 3841 GCAAGTATTACCAACCAACAGCTGATATACAGCGGTAGATTTCCGGTAAACACGTTAAGT 3900  
Db 3841 GCAAGTATTACCAACCAACAGCTGATATACAGCGGTAGATTTCCGGTAAACACGTTAAGT 3900  
QY 3901 GTTAGCGGACTGGTGAATTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960  
Db 3901 GTTAGCGGACTGGTGAATTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGGT 3960  
QY 3961 GAGGCTAATTAACAAAGTGAACAGGTACAATTTGCGGTACAATTTCCGGTAAATACGGTA 4020  
Db 3961 GAGGCTAATTAACAAAGTGAACAGGTACAATTTGCGGTACAATTTCCGGTAAATACGGTA 4020  
QY 4021 AATGTTACGCAACAGCTGGCGATTTACAGTTGGGAATGCGCAGAAATTAATGCGACA 4080  
Db 4021 AATGTTACGCAACAGCTGGCGATTTACAGTTGGGAATGCGCAGAAATTAATGCGACA 4080  
QY 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGAAATACCTTACTACTGAAGCGGTCTAGC 4140  
Db 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGAAATACCTTACTACTGAAGCGGTCTAGC 4140  
QY 4141 ATCACTTCACTAAGGCTCAGGTAGACCTCTTGCTCAGAAATGAGTACGAGGAGG 4200  
Db 4141 ATCACTTCACTAAGGCTCAGGTAGACCTCTTGCTCAGAAATGAGTACGAGGAGG 4200  
QY 4201 ATTAATGCTGCTAATGACATTAATACTACAGGCAACCTTAACCGGTCGAGGCTCG 4260  
Db 4201 ATTAATGCTGCTAATGACATTAATACTACAGGCAACCTTAACCGGTCGAGGCTCG 4260  
QY 4261 GATATTAAGCAACCGGCTGCTGTTATTAACGCAAAAGATGCTAAGCTAAATGCT 4320  
Db 4261 GATATTAAGCAACCGGCTGCTGTTATTAACGCAAAAGATGCTAAGCTAAATGCT 4320

QY 4321 GATGCATCAGGTGATAGTACAGAAAGTCAATGCAAGCAAGCGCTCTGCTAGTGTG 4380  
Db 4321 GATGCATCAGGTGATAGTACAGAAAGTCAATGCAAGCAAGCGCTCTGCTAGTGTG 4380  
QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATACATCACTGGGGATTTAAACACAGTAATGGTTA 4440  
Db 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATACATCACTGGGGATTTAAACACAGTAATGGTTA 4440  
QY 4441 AATATCATTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGAAATGAGGTG 4500  
Db 4441 AATATCATTTCCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGAAATGAGGTG 4500  
QY 4501 AATATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGCAACCGCTCTT 4560  
Db 4501 AATATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTAAATTTGAAGCGCAACCGCTCTT 4560  
QY 4561 GAAAGATTAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTTAACTTTGGTGAAGT 4620  
Db 4561 GAAAGATTAAGATTTATCTGATGAAGAAAGAGAAACATTAAGCTTAACTTTGGTGAAGT 4620  
QY 4621 GCTGTACGTTTGTGTTGAGCCAAATTAATACAAATTTACAGTCAATACACAATTAATTA 4680  
Db 4621 GCTGTACGTTTGTGTTGAGCCAAATTAATACAAATTTACAGTCAATACACAATTAATTA 4680  
QY 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTCTTCTCAAGTGAAGT 4740  
Db 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGTAAAGCGGTCTTCTCAAGTGAAGT 4740  
QY 4741 GCGGCAACAGTATGTACCAATTTGCTGACGATGCAAGCGGTCTTCTCAAGTGAAGT 4800  
Db 4741 GCGGCAACAGTATGTACCAATTTGCTGACGATGCAAGCGGTCTTCTCAAGTGAAGT 4800  
QY 4801 GTAGATTTCACTCCGCAATGAAGTCAATTTTCTGATTAATTTACTGTTGCGGTTAA 4860  
Db 4801 GTAGATTTCACTCCGCAATGAAGTCAATTTTCTGATTAATTTACTGTTGCGGTTAA 4860  
QY 4861 GTTCAGTACGCGCTTACCCTCTTTGTAATAATTTACGGAATTAATTAAGTATTTT 4920  
Db 4861 GTTCAGTACGCGCTTACCCTCTTTGTAATAATTTACGGAATTAATTAAGTATTTT 4920  
QY 4921 AACAGGTTATTTATG 4937  
Db 4921 AACAGGTTATTTATG 4937

## RESULT 6

US-08-617-697-3

Sequence 3, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,697

FILING DATE: 01-APR-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:



APPLICATION NUMBER: US 08/302,832  
 FILING DATE: 05-OCT-1994  
 PRIOR APPLICATION DATA: PCT/US93/02166  
 APPLICATION NUMBER: -1993  
 FILING DATE: 16-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 BEST COPY W  
 REGISTRATION NUMBER: 22,551  
 REFERENCE/DOCKET NUMBER: 1038-557  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4937 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-617-697-3

Query Match	100.0%	Score 4937	DB 2	Length 4937	0
Best Local Similarity	100.0%	Pred. No. 0			
Matches 4937	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1	TAATATACAGATATATAAAATAAATCAAGATTTTGTGATGACAAACAAANTACAA	60		
DB	1	TAATATACAGATATATAAAATAAATCAAGATTTTGTGATGACAAACAAANTACAA	60		
QY	61	CACCTTTTTCGAGTCTATATGCAAAATATTTTAAAAAAATAGTATAAATCCGCCATATAA	120		
DB	61	CACCTTTTTCGAGTCTATATGCAAAATATTTTAAAAAAATAGTATAAATCCGCCATATAA	120		
QY	121	AATGGTATAATCTTTCATCTTTTCATCTTTTAATCTTTTCATCTTTTCATCTTTTCAT	180		
DB	121	AATGGTATAATCTTTCATCTTTTCATCTTTTAATCTTTTCATCTTTTCATCTTTTCAT	180		
QY	181	CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT	240		
DB	181	CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT	240		
QY	241	CACATGAATGATGAACCGAGGGAAGGAGGAGGGGCAAGATGAAGAGGGAGCTGAAC	300		
DB	241	CACATGAATGATGAACCGAGGGAAGGAGGAGGGGCAAGATGAAGAGGGAGCTGAAC	300		
QY	301	GAACGCAATGATGAAGTAATTTAATCTTCACTTAACCTTAGCAGAAATATGAACAAG	360		
DB	301	GAACGCAATGATGAAGTAATTTAATCTTCACTTAACCTTAGCAGAAATATGAACAAG	360		
QY	361	ATATATCGTCTCAAAATTCAGCAAAACGCTTGATGCTTTGGTTGCTGTGCTGATATGGCA	420		
DB	361	ATATATCGTCTCAAAATTCAGCAAAACGCTTGATGCTTTGGTTGCTGTGCTGATATGGCA	420		
QY	421	CGGGGTGTGACCATTCACAGAAAAGGCTTCCGCTATGTTACTATCTTTTAGGTGAAC	480		
DB	421	CGGGGTGTGACCATTCACAGAAAAGGCTTCCGCTATGTTACTATCTTTTAGGTGAAC	480		
QY	481	CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATTAATCCA	540		
DB	481	CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATTAATCCA	540		
QY	541	CAATCTCTTTTAGCAAGCGGCTTTACAAGGAATGGATGTAGTACAGGCACAGCCACTATG	600		
DB	541	CAATCTCTTTTAGCAAGCGGCTTTACAAGGAATGGATGTAGTACAGGCACAGCCACTATG	600		
QY	601	CAAGTAGATGGTAAATAAAACCATTTCCCGCAACAGTGTGACGCTATCATTAATTTGGAAA	660		
DB	601	CAAGTAGATGGTAAATAAAACCATTTCCCGCAACAGTGTGACGCTATCATTAATTTGGAAA	660		
QY	661	CAATTTAACATCGACCAAAATGAATGGTGACGTTTTTACAGAAACAAACAACCTCCGCC	720		
DB	661	CAATTTAACATCGACCAAAATGAATGGTGACGTTTTTACAGAAACAAACAACCTCCGCC	720		



Db 1801 ACCAATACAACCTATTTCAAAATATCTGAAAAGCGCTGACCAATGATATACGGCATCA 1860  
QY 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCOCAT 1920  
Db 1861 AGAAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTCOCAT 1920  
QY 1921 AGTAAAGCTACGCTGGGAGGCGTTCAGATTGATGAGATATTAATCTTAAGGGCGA 1980  
Db 1921 AGTAAAGCTACGCTGGGAGGCGTTCAGATTGATGAGATATTAATCTTAAGGGCGA 1980  
QY 1981 AATTTAAACCAATTAATCTCGGGGATGGTTGATGTTTCAATAAATATTAACGCTTGATCAG 2040  
Db 1981 AATTTAAACCAATTAATCTCGGGGATGGTTGATGTTTCAATAAATATTAACGCTTGATCAG 2040  
QY 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATAAACAAGCAAGC 2100  
Db 2041 GGTTTTTTAAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATAAACAAGCAAGC 2100  
QY 2101 GACGCGCAAAATGCTTAAATTTGTCGCCAGGCACTGTAAACATTTACAGGAGGGAAA 2160  
Db 2101 GACGCGCAAAATGCTTAAATTTGTCGCCAGGCACTGTAAACATTTACAGGAGGGAAA 2160  
QY 2161 GATTTACAGGCTTACCAACGCTATCTTTAAACGGAACGGTAAAGGCTGAATATCATTTCA 2220  
Db 2161 GATTTACAGGCTTACCAACGCTATCTTTAAACGGAACGGTAAAGGCTGAATATCATTTCA 2220  
QY 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280  
Db 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGGCACAAATTAACATATCTGGGAATATAACA 2280  
QY 2281 ATTAACCAAACTACGAGAAAGAACACCTCTGATTTGGCAACACGAGCATGATCGCACTGG 2340  
Db 2281 ATTAACCAAACTACGAGAAAGAACACCTCTGATTTGGCAACACGAGCATGATCGCACTGG 2340  
QY 2341 AAGCTCAGTGTCTTAATCTAGAGACAGGCGCAATTTTACCTTTTAAATACATTTCA 2400  
Db 2341 AAGCTCAGTGTCTTAATCTAGAGACAGGCGCAATTTTACCTTTTAAATACATTTCA 2400  
QY 2401 AGCAATAGCAAAAGCTTTAACAACACAGTATAGAAGCTCTGACGGGTGCAATTTAACGGC 2460  
Db 2401 AGCAATAGCAAAAGCTTTAACAACACAGTATAGAAGCTCTGACGGGTGCAATTTAACGGC 2460  
QY 2461 GTAATGCAACATGTCATCTCAATCTCAAGAGAGGAGGAAAGTTAATTTCAAAATFAAA 2520  
Db 2461 GTAATGCAACATGTCATCTCAATCTCAAGAGAGGAGGAAAGTTAATTTCAAAATFAAA 2520  
QY 2521 CCAACAGAAACATGAACACAAAGCACTTTACCAATTCGGTTTTTAGCCAATATCA 2580  
Db 2521 CCAACAGAAACATGAACACAAAGCACTTTACCAATTCGGTTTTTAGCCAATATCA 2580  
QY 2581 GCCACTGTGGGGCTCTGTTTTTTTTTTTATATATATGCAACCACTTTAGCCAATATCA 2640  
Db 2581 GCCACTGTGGGGCTCTGTTTTTTTTTTTATATATATGCAACCACTTTAGCCAATATCA 2640  
QY 2641 GAGTTAAAAATGAGTGAAATTAATATCTTAACGGCGCTTAATTTTACCTTAATTTCCCAT 2700  
Db 2641 GAGTTAAAAATGAGTGAAATTAATATCTTAACGGCGCTTAATTTTACCTTAATTTCCCAT 2700  
QY 2701 GTTTCGGGCGCATGACGCTTTTAAATCAACAAGACTTAACCAATAAATGCAACCAATTC 2760  
Db 2701 GTTTCGGGCGCATGACGCTTTTAAATCAACAAGACTTAACCAATAAATGCAACCAATTC 2760  
QY 2761 AATTTACGCTCAGACAGCAAGAGATGATTTTTTATGAGGGTACGAGGCAATGCCATC 2820  
Db 2761 AATTTACGCTCAGACAGCAAGAGATGATTTTTTATGAGGGTACGAGGCAATGCCATC 2820  
QY 2821 AATTTACACCTTACACATATCCATCTGGGGTAAATGTCACCCCTTGGTGCAAAACTCA 2880  
Db 2821 AATTTACACCTTACACATATCCATCTGGGGTAAATGTCACCCCTTGGTGCAAAACTCA 2880  
QY 2881 AGAGAGCAATTAACGGGAATTAATTAATCTGAGAAAGCAGCAATTTACGCTAGAAAGCC 2940  
Db 2881 AGAGAGCAATTAACGGGAATTAATTAATCTGAGAAAGCAGCAATTTACGCTAGAAAGCC 2940

Db 2881 AGCAGCAGCAATTAACGGGAATTAATTAATCTGAGAAAGCAGCAATTTACGCTAGAAAGCC 2940  
QY 2941 AATTAACGCCCTTAATTAACGCAAAACATTAAGGATAGAGTTAATAAATTTGGCAGCTGCTC 3000  
Db 2941 AATTAACGCCCTTAATTAACGCAAAACATTAAGGATAGAGTTAATAAATTTGGCAGCTGCTC 3000  
QY 3001 GTTAATGGAGTTTAAGTTTAACGCGCAAAATGCAGATATTAAGGCAATCTCACTATT 3060  
Db 3001 GTTAATGGAGTTTAAGTTTAACGCGCAAAATGCAGATATTAAGGCAATCTCACTATT 3060  
QY 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCATAATATCACCGCAATTTT 3120  
Db 3061 TCAGAAAGCGCCACTTTTAAAGGAAAGACTAGAGATACCCATAATATCACCGCAATTTT 3120  
QY 3121 ACCAATTAATGSCACTGCGGCAATTAATTAACAAGAGTGGTAAAACTTTGGCAATGTT 3180  
Db 3121 ACCAATTAATGSCACTGCGGCAATTAATTAACAAGAGTGGTAAAACTTTGGCAATGTT 3180  
QY 3181 ACCAATTAATGSGTATTTAAACATTAACCACTCACCTAAAGCAAGAGCATCATC 3240  
Db 3181 ACCAATTAATGSGTATTTAAACATTAACCACTCACCTAAAGCAAGAGCATCATC 3240  
QY 3241 GCGGAGATATAATCAACAAGAGAGCTTAATAATTAACAGACAGTAATATGATGCT 3300  
Db 3241 GCGGAGATATAATCAACAAGAGAGCTTAATAATTAACAGACAGTAATATGATGCT 3300  
QY 3301 GAAATCCAAATTTGGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTTCCGAT 3360  
Db 3301 GAAATCCAAATTTGGGCGCAATATCTCGCAAAAGAGGCAACCTCACGATTTTCCGAT 3360  
QY 3361 AAAATTAATATCAACCAACAGATAACATCAAAAGGTTTATGAGGAGGACTCTAGT 3420  
Db 3361 AAAATTAATATCAACCAACAGATAACATCAAAAGGTTTATGAGGAGGACTCTAGT 3420  
QY 3421 TCAGATGCGCAAGTAAATGCCAACCCTAATTTAAACCAAGAAATTAATGACAGAA 3480  
Db 3421 TCAGATGCGCAAGTAAATGCCAACCCTAATTTAAACCAAGAAATTAATGACAGAA 3480  
QY 3481 GACCTAAGTATTTCAAGTTTCAATTAAGCAGAGATTACAGCCAAAGATGTTAGAGATTTA 3540  
Db 3481 GACCTAAGTATTTCAAGTTTCAATTAAGCAGAGATTACAGCCAAAGATGTTAGAGATTTA 3540  
QY 3541 ACTATTGCAACAGTAAATGCGGTAAACGCGTTCGCAAGGCGGTAACAGCAATTTAAC 3600  
Db 3541 ACTATTGCAACAGTAAATGCGGTAAACGCGTTCGCAAGGCGGTAACAGCAATTTAAC 3600  
QY 3601 AATGTTAAGATTTCAAAATCTCTGCTGACGGTCACAATGTGACACTTAATAGCAAGTG 3660  
Db 3601 AATGTTAAGATTTCAAAATCTCTGCTGACGGTCACAATGTGACACTTAATAGCAAGTG 3660  
QY 3661 AAAACATCTAGCAGCAATGCGGACGTGAAGCAATAGCGACAGTACCGGCTTAAC 3720  
Db 3661 AAAACATCTAGCAGCAATGCGGACGTGAAGCAATAGCGACAGTACCGGCTTAAC 3720  
QY 3721 ATTACTGCAAAATGTAAGAGTAAACAAGATATTAATCTCTCAAAACAGTAATATC 3780  
Db 3721 ATTACTGCAAAATGTAAGAGTAAACAAGATATTAATCTCTCAAAACAGTAATATC 3780  
QY 3781 ACCCGCTGCGAAAGGTTTACCACACAGGCGCTGACCATTAACGCAACAAATGGCA 3840  
Db 3781 ACCCGCTGCGAAAGGTTTACCACACAGGCGCTGACCATTAACGCAACAAATGGCA 3840  
QY 3841 GCAAGTATTACCAACCAAGAGTATATCAGCGGTACGATTTCCGCTAACACGTAAGT 3900  
Db 3841 GCAAGTATTACCAACCAAGAGTATATCAGCGGTACGATTTCCGCTAACACGTAAGT 3900  
QY 3901 GTTAGCGGACTGCTGATTTTAAACCACTTAATCCGCTCAAAATTTGAACGGAATTCGGT 3960  
Db 3901 GTTAGCGGACTGCTGATTTTAAACCACTTAATCCGCTCAAAATTTGAACGGAATTCGGT 3960  
QY 3961 GAGGCTAATGTAACAGTGAACAGGTACAATTTGGCGGTACAATTTCCGCTAACACGTA 4020  
Db 3961 GAGGCTAATGTAACAGTGAACAGGTACAATTTGGCGGTACAATTTCCGCTAACACGTA 4020

TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd.  
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 STREET: Bldg 1  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/719,641  
 FILING DATE:

CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/302,832  
 FILING DATE: 16-SEP-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US93/02166  
 FILING DATE: 16-MAR-1993

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9205704.1  
 FILING DATE: 16-MAR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Berks & Berks, J. & W.  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: 1038-625  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0810

INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4937 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-719-641-3

Query Match 100.0%; Score 4937; DB 4; Length 4937;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4021 AATGTTACGGCAACCGCTGGGATTTACAGCTTGGGAATGCGCAGAAATTAATGCGACA 4080  
 DB 4021 AATGTTACGGCAACCGCTGGGATTTACAGCTTGGGAATGCGCAGAAATTAATGCGACA 4080  
 QY 4081 GAAGAGCTGCAACCTTTAACGCCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140  
 DB 4081 GAAGAGCTGCAACCTTTAACGCCAACAGGGAATACCTTGACTACTGAAGCGGTTCTAGC 4140  
 QY 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATGCGAGGAAGC 4200  
 DB 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGAATGGTAGCATGCGAGGAAGC 4200  
 QY 4201 ATTAATGCTGCTAAATGTGACATTAATACTACAGGCACCTTAACCCAGCTGGCAGGCTCG 4260  
 DB 4201 ATTAATGCTGCTAAATGTGACATTAATACTACAGGCACCTTAACCCAGCTGGCAGGCTCG 4260  
 QY 4261 GATATTAAGCAACACCGCGACCTTGGTTATTACGCAAAAGATGCTAAAGCTAAATGGT 4320  
 DB 4261 GATATTAAGCAACACCGCGACCTTGGTTATTACGCAAAAGATGCTAAAGCTAAATGGT 4320  
 QY 4321 GATGCAATCAGGTGATAGTACAGAAGTGAATGAGTCAAGCCAGCGGCTCTGGTAGTGTG 4380  
 DB 4321 GATGCAATCAGGTGATAGTACAGAAGTGAATGAGTCAAGCCAGCGGCTCTGGTAGTGTG 4380  
 QY 4381 ACTGCGCAACCTTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440  
 DB 4381 ACTGCGCAACCTTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAAATGGGTTA 4440  
 QY 4441 AATATATCTTTCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATGAGGTG 4500  
 DB 4441 AATATATCTTTCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATGAGGTG 4500  
 QY 4501 AATATATCTTTCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATGAGGTG 4560  
 DB 4501 AATATATCTTTCGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATGAGGTG 4560  
 QY 4561 GAAAGCTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACTTGGTGTAACT 4620  
 DB 4561 GAAAGCTAAAGATTTATCTGATGAAGAAAGAGAAACATTTAGCTAAACTTGGTGTAACT 4620  
 QY 4621 GCTGAGCTTTGCTGAGCCAAATATACAAATTTACAGTCAATACACAAAATGAATTTACA 4680  
 DB 4621 GCTGAGCTTTGCTGAGCCAAATATACAAATTTACAGTCAATACACAAAATGAATTTACA 4680  
 QY 4681 ACCAGACCGTCAAGTCAAGTCAATTTCTGAGGTAAGCGGTTTCTCAAGTGTGTAAT 4740  
 DB 4681 ACCAGACCGTCAAGTCAAGTCAATTTCTGAGGTAAGCGGTTTCTCAAGTGTGTAAT 4740  
 QY 4741 GCGGACGAGTATGTACCAATGTTGCTGAGCATGGACAGCGGTAGTCAGTAATGACAG 4800  
 DB 4741 GCGGACGAGTATGTACCAATGTTGCTGAGCATGGACAGCGGTAGTCAGTAATGACAG 4800  
 QY 4801 GTAGATTTTCATCCCTGCAATGAAGTCATTTTATTTTCGTATTTTACTGTGGGTTAAA 4860  
 DB 4801 GTAGATTTTCATCCCTGCAATGAAGTCATTTTATTTTCGTATTTTACTGTGGGTTAAA 4860  
 QY 4861 GTTCAGTACGGCTTTACCCATCTTGTAAAAAATTTACGGAGATACAAATAAAGTATTTT 4920  
 DB 4861 GTTCAGTACGGCTTTACCCATCTTGTAAAAAATTTACGGAGATACAAATAAAGTATTTT 4920  
 QY 4921 AACAGGTTATTTATG 4937  
 DB 4921 AACAGGTTATTTATG 4937

RESULT 7  
 US-08-719-641-3  
 Sequence 3, Application US/08719641  
 Patent No. 6218141  
 GENERAL INFORMATION:  
 APPLICANT: Barenkamp, Stephen J  
 TITLE OF INVENTION: High Molecular Weight Surface Proteins

|||||  
Db 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAAATGCTTTGGTGTGTGCTGAATGGCA 420  
QY 421 CGGGGTTGTGACCAATTCACAGAAAAAGGCTTCGGCTATGTTACTATCTTTAGTGTAAAC 480  
Db 421 CGGGGTTGTGACCAATTCACAGAAAAAGGCTTCGGCTATGTTACTATCTTTAGTGTAAAC 480  
QY 481 CACTTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGTAAACATCTATTCCA 540  
Db 481 CACTTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGTGTAAACATCTATTCCA 540  
QY 541 CAATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGTAGTACACAGCCACAGCCACTATG 600  
Db 541 CAATCTGTTTTAGCAAGCGGCTTACAAGGAATGGATGTAGTACACAGCCACAGCCACTATG 600  
QY 601 CAAGTAGATGGTAAATAAACCAATTTACCAACAGTGTTCACGCTATCAATTAATTTGAAA 660  
Db 601 CAAGTAGATGGTAAATAAACCAATTTACCAACAGTGTTCACGCTATCAATTAATTTGAAA 660  
QY 661 CAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTTACAAGAAAACAACATCCGCC 720  
Db 661 CAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTTACAAGAAAACAACATCCGCC 720  
QY 721 GTATTTCAACCGGTTTACATCTAACCAAAATTCCTCCCAATTTAAAGGGATTTAGATTTCTAAC 780  
Db 721 GTATTTCAACCGGTTTACATCTAACCAAAATTCCTCCCAATTTAAAGGGATTTAGATTTCTAAC 780  
QY 781 GGACAGTCTTTTTTAATCAACCAAAATGGTATCACAAATAGTAAAGAGCAATTAATTAAC 840  
Db 781 GGACAGTCTTTTTTAATCAACCAAAATGGTATCACAAATAGTAAAGAGCAATTAATTAAC 840  
QY 841 ACTAATGGCTTTACGGCTCTACGCTAGACATTTCTAACCAAAATGGTAAAGAGCAATTTAAAT 900  
Db 841 ACTAATGGCTTTACGGCTCTACGCTAGACATTTCTAACCAAAATGGTAAAGAGCAATTTAAAT 900  
QY 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCAGCGTTAAAT 960  
Db 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCAGCGTTAAAT 960  
QY 961 ACTGTCGTTAAAGACGGCAGTGTAAATCTTAATGGTGGCAAAATGCAAAAGAGGCTGTG 1020  
Db 961 ACTGTCGTTAAAGACGGCAGTGTAAATCTTAATGGTGGCAAAATGCAAAAGAGGCTGTG 1020  
QY 1021 ATTACGTTAAATGGTGGCAGCAATTTCTTACTCGAGGCAAAATTAACCATCAGCGAT 1080  
Db 1021 ATTACGTTAAATGGTGGCAGCAATTTCTTACTCGAGGCAAAATTAACCATCAGCGAT 1080  
QY 1081 ATAATAAACCCCAACCAATTTACTTACAGCAATTCGCGCTGAAATGCAAGCGTCAATCTG 1140  
Db 1081 ATAATAAACCCCAACCAATTTACTTACAGCAATTCGCGCTGAAATGCAAGCGTCAATCTG 1140  
QY 1141 GCGATATTTTCCCAAGCGGTAACATTAATGTCGCTGCGCACTATTTCGCAACCAA 1200  
Db 1141 GCGATATTTTCCCAAGCGGTAACATTAATGTCGCTGCGCACTATTTCGCAACCAA 1200  
QY 1201 GGTAACTTTCTGCTGATTTGTAAGCAAAAGATAAAGCGCAATATTGCTCTTCCTCCGC 1260  
Db 1201 GGTAACTTTCTGCTGATTTGTAAGCAAAAGATAAAGCGCAATATTGCTCTTCCTCCGC 1260  
QY 1261 AAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320  
Db 1261 AAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320  
QY 1321 GCGAGCTGATGATTAACAGCGGATAAAGTACATTAATAACAGGTGCGAGTTATCGACCTT 1380  
Db 1321 GCGAGCTGATGATTAACAGCGGATAAAGTACATTAATAACAGGTGCGAGTTATCGACCTT 1380  
QY 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTGGCGTGACAGCGCGCGGAGGTAATAAC 1440  
Db 1381 TCAGGTAAAGAGGGGAGAACTTTACCTTGGCGTGACAGCGCGCGGAGGTAATAAC 1440  
QY 1441 GCGATTAATTAAGCAAAAGCACTTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500  
|||||

Db 1441 GGCATTCAATTTAGCAAAAGAAACCCCTTTAGAAAAAGGCTCAACCATCAATGATCAGGC 1500  
QY 1501 AAAGAAAAAGGCGGACGCGCTATTGTTGGGCGGATATTGCGTTAAATTCACCGCAATATT 1560  
Db 1501 AAAGAAAAAGGCGGACGCGCTATTGTTGGGCGGATATTGCGTTAAATTCACCGCAATATT 1560  
QY 1561 AACGCTCAAGGTAGTGTGATATCGCTAAACCCGGTGGTTTTGTTGGAGACATCGGGGCAT 1620  
Db 1561 AACGCTCAAGGTAGTGTGATATCGCTAAACCCGGTGGTTTTGTTGGAGACATCGGGGCAT 1620  
QY 1621 TATTTATCCATTGACAGCAATGCAATTTTAAACAAAAAGAGTGTGTTAGACCCCTGAT 1680  
Db 1621 TATTTATCCATTGACAGCAATGCAATTTTAAACAAAAAGAGTGTGTTAGACCCCTGAT 1680  
QY 1681 GATGTAACAAATTTGAAGCGGAGACCCCTTCGCAATTAATACCGGTATTAATGATGAATTC 1740  
Db 1681 GATGTAACAAATTTGAAGCGGAGACCCCTTCGCAATTAATACCGGTATTAATGATGAATTC 1740  
QY 1741 CCAACAGGACCGGTGAAGCAAGCGACCCCTAAAAAATAGCGAACTCAAAACACCGCTA 1800  
Db 1741 CCAACAGGACCGGTGAAGCAAGCGACCCCTAAAAAATAGCGAACTCAAAACACCGCTA 1800  
QY 1801 ACCAATACAACTATTTTCAAAATTTATCTGAAAAACGCTTGAACAACTCAAAACACCGCTA 1860  
Db 1801 ACCAATACAACTATTTTCAAAATTTATCTGAAAAACGCTTGAACAACTCAAAACACCGCTA 1860  
QY 1861 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTTCTCCAT 1920  
Db 1861 AGAAACCTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTTCTCCAT 1920  
QY 1921 AGTAAAGGTACGCGGAGCGGCTTCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980  
Db 1921 AGTAAAGGTACGCGGAGCGGCTTCAGATTTGATGGAGATATTACTTCTAAAGCGGA 1980  
QY 1981 AATTTAAACCAATTTATTTCTGCGGAGTGGTTGATGTTCTATAAAATATTACGCTTGATCAG 2040  
Db 1981 AATTTAAACCAATTTATTTCTGCGGAGTGGTTGATGTTCTATAAAATATTACGCTTGATCAG 2040  
QY 2041 GGTTTTTTAATATTACCCTCCCTTTCCGTAGCTTTTGAAGGTGGAATTAACAAGCACGC 2100  
Db 2041 GGTTTTTTAATATTACCCTCCCTTTCCGTAGCTTTTGAAGGTGGAATTAACAAGCACGC 2100  
QY 2101 GACGCGCAATCTAAATTTGTCGCCAGGCGACTGTAAACCATTAACAAGCACGC 2160  
Db 2101 GACGCGCAATCTAAATTTGTCGCCAGGCGACTGTAAACCATTAACAAGCACGC 2160  
QY 2161 GATTTACGGGCTAACACGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCAATCA 2220  
Db 2161 GATTTACGGGCTAACACGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCAATCA 2220  
QY 2221 TCAGTGAATTAATTTAACCCCAATCTTAGTGCGCAATTAACATATCTGGGAATATAACA 2280  
Db 2221 TCAGTGAATTAATTTAACCCCAATCTTAGTGCGCAATTAACATATCTGGGAATATAACA 2280  
QY 2281 ATTAAACCAACTACAGAAAGAACACCTCGTATTGGCAAAACGCGGATGATTCGACCTGG 2340  
Db 2281 ATTAAACCAACTACAGAAAGAACACCTCGTATTGGCAAAACGCGGATGATTCGACCTGG 2340  
QY 2341 AACGCTAGTCTCTTAATCTAGACAGCGGCAATTTTACCTTTTAAATACATTTCA 2400  
Db 2341 AACGCTAGTCTCTTAATCTAGACAGCGGCAATTTTACCTTTTAAATACATTTCA 2400  
QY 2401 AGCAATAGCAAGGCTTTAAACACACAGTATAGAAGCTCTGACGGGTGAATTTTAAAGGC 2460  
Db 2401 AGCAATAGCAAGGCTTTAAACACACAGTATAGAAGCTCTGACGGGTGAATTTTAAAGGC 2460  
QY 2461 GTAAATGGCAACATCTCAATCTCAAAAGAGGAGGCAAGCTTAATTTCAATTTAAA 2520  
Db 2461 GTAAATGGCAACATCTCAATCTCAAAAGAGGAGGCAAGCTTAATTTCAATTTAAA 2520  
QY 2521 CCAACGAGCAACATGACACACAAACCTTTTACCCTTTTGGTTTTTGGCAATATCACA 2580  
Db 2521 CCAACGAGCAACATGACACACAAACCTTTTACCCTTTTGGTTTTTGGCAATATCACA 2580

QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGGATATATATGCGAACCAATCTGCGCAGAGGGCT 2640  
 DB 2581 GCCACTGGTGGGGCTCTGTTTTTTTGGATATATATGCGAACCAATCTGCGCAGAGGGCT 2640  
 QY 2641 GAGTTAAAAAAGTGAATAATATCTCTAACGGCGCTAAATTTTACCTTTAAATCCCAT 2700  
 DB 2641 GAGTTAAAAAAGTGAATAATATCTCTAACGGCGCTAAATTTTACCTTTAAATCCCAT 2700  
 QY 2701 GTTTCGGCGGATGAGCGCTTTTAAAAATCAACAAAGACTTTAACCATATAATGCAACCAATFCA 2760  
 DB 2701 GTTTCGGCGGATGAGCGCTTTTAAAAATCAACAAAGACTTTAACCATATAATGCAACCAATFCA 2760  
 QY 2761 AATTTTCAGCCTCAGACAGACGAAAGATGATTTTTATGACGGGTACGCGCAATGCCATC 2820  
 DB 2761 AATTTTCAGCCTCAGACAGACGAAAGATGATTTTTATGACGGGTACGCGCAATGCCATC 2820  
 QY 2821 AATTTCAACCTACCAATATCCATCTTGGCGGTAATGTCAACCTTTGGTGGACAAACTCA 2880  
 DB 2821 AATTTCAACCTACCAATATCCATCTTGGCGGTAATGTCAACCTTTGGTGGACAAACTCA 2880  
 QY 2881 ACCAGCAGCATTTACGGGGAATTTACTATCGAGAAAGACGCAAAATGTTACGCTAGAAGCC 2940  
 DB 2881 ACCAGCAGCATTTACGGGGAATTTACTATCGAGAAAGACGCAAAATGTTACGCTAGAAGCC 2940  
 QY 2941 AATAAGCGCCCTAATCAGCAAAACATTAAGGATAGAGTTATAAATCTTGGCAGCTTGCTC 3000  
 DB 2941 AATAAGCGCCCTAATCAGCAAAACATTAAGGATAGAGTTATAAATCTTGGCAGCTTGCTC 3000  
 QY 3001 GTTAATGGGAGTTAACTGGCGGAAATGCGAGATATTAAGGCAATCTCACTATT 3060  
 DB 3001 GTTAATGGGAGTTAACTGGCGGAAATGCGAGATATTAAGGCAATCTCACTATT 3060  
 QY 3061 TCAGAAAGCCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATCACCGGCAATTTT 3120  
 DB 3061 TCAGAAAGCCCACTTTTAAAGGAAAGACTAGAGATACCTTAATATCACCGGCAATTTT 3120  
 QY 3121 ACCAATATGCGATGCGGCAATTAATATACACAGAGAGTGTGTAACCTTGGCAATGTT 3180  
 DB 3121 ACCAATATGCGATGCGGCAATTAATATACACAGAGAGTGTGTAACCTTGGCAATGTT 3180  
 QY 3181 ACCAATGATGGTATTTAAACATTTACCCTCACGCTTAACGCAACCAAGAGCATCATC 3240  
 DB 3181 ACCAATGATGGTATTTAAACATTTACCCTCACGCTTAACGCAACCAAGAGCATCATC 3240  
 QY 3241 GCGGAGATATAATCAACAAAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300  
 DB 3241 GCGGAGATATAATCAACAAAAAGGAGCTTAAATATTACAGACAGTAAATATGATGCT 3300  
 QY 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAAGAGGCAACCTTACGATTTCTTCGGAT 3360  
 DB 3301 GAAATCCAAATTTGGCGGCAATATCTCGCAAAAAGAGGCAACCTTACGATTTCTTCGGAT 3360  
 QY 3361 AAAATTAATATCACCACACATTAACATCAAAAGGATTTGATGGAGAGGACTCTAGT 3420  
 DB 3361 AAAATTAATATCACCACACATTAACATCAAAAGGATTTGATGGAGAGGACTCTAGT 3420  
 QY 3421 TCAGATGGACAGTAATGCAACCTTAATTAACCAAGGATTTGAAATTCACAGAA 3480  
 DB 3421 TCAGATGGACAGTAATGCAACCTTAATTAACCAAGGATTTGAAATTCACAGAA 3480  
 QY 3481 GACCTAAGTATTTACAGTTTCAATTAACAGCAGATTAACGCAAGAGTGTAGAGATTTA 3540  
 DB 3481 GACCTAAGTATTTACAGTTTCAATTAACAGCAGATTAACGCAAGAGTGTAGAGATTTA 3540  
 QY 3541 ACTATTGCAACAGTAATGACCGGTAAACAGCGGTGCGGAGCGGCAACAGTAACTTTAAC 3600  
 DB 3541 ACTATTGCAACAGTAATGACCGGTAAACAGCGGTGCGGAGCGGCAACAGTAACTTTAAC 3600  
 QY 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGCTCAAAATGTCACACTAAATAGCAAGTG 3660  
 DB 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGCTCAAAATGTCACACTAAATAGCAAGTG 3660

QY 3661 AAAACATCTAGCAGCAATGCGGACGCTGAAGCAATAGCGACAAACGATACCGGCTTAACT 3720  
 DB 3661 AAAACATCTAGCAGCAATGCGGACGCTGAAGCAATAGCGACAAACGATACCGGCTTAACT 3720  
 QY 3721 ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAAACAGTAAATATC 3780  
 DB 3721 ATTACTGCAAAAAATGTAGAAGTAAACAAAGATATTACTTCTCTCAAAAACAGTAAATATC 3780  
 QY 3781 ACCCGCTCGGAAAGGTTTACCACACAGCAGGCTCGACCAATTAACGCAACAAATGCGCAA 3840  
 DB 3781 ACCCGCTCGGAAAGGTTTACCACACAGCAGGCTCGACCAATTAACGCAACAAATGCGCAA 3840  
 QY 3841 GCAAGTATTACACCAACCAACAGGTGATCAGCGGTACGATTTCCGGTACACAGGTAACT 3900  
 DB 3841 GCAAGTATTACACCAACCAACAGGTGATCAGCGGTACGATTTCCGGTAAACAGGTAACT 3900  
 QY 3901 GTTACGGGACTGTGATTTTAAACCACTAAATCCCGCTCAAAAATTTGAAGCAAAATCCGGT 3960  
 DB 3901 GTTACGGGACTGTGATTTTAAACCACTAAATCCCGCTCAAAAATTTGAAGCAAAATCCGGT 3960  
 QY 3961 GAGGCTAATGTAAACAGGTGCAACAGGTACAAATTCGGGTACAAATTTCCGGTAAATACGTA 4020  
 DB 3961 GAGGCTAATGTAAACAGGTGCAACAGGTACAAATTTCCGGTACAAATTTCCGGTAAATACGTA 4020  
 QY 4021 AATGTTACGGCAACGCTGGCGATTTTAAACAGTTGGGAATGCGGCAAAATTTAAATGCGNCA 4080  
 DB 4021 AATGTTACGGCAACGCTGGCGATTTTAAACAGTTGGGAATGCGGCAAAATTTAAATGCGNCA 4080  
 QY 4081 GAAAGGAGCTGCAACCTTAAACGCAACAGGGAATACCTTTGACTACTGAAAGCGGTTCTAGC 4140  
 DB 4081 GAAAGGAGCTGCAACCTTAAACGCAACAGGGAATACCTTTGACTACTGAAAGCGGTTCTAGC 4140  
 QY 4141 ATCACTTCAACTTAAGGGTCAAGTAGACCTTTGGCTCAGAAATGAGTATGCGAAGG 4200  
 DB 4141 ATCACTTCAACTTAAGGGTCAAGTAGACCTTTGGCTCAGAAATGAGTATGCGAAGG 4200  
 QY 4201 ATTAATGCTGCTTAATGTGACATTAATACAGGCACTTTAAACCAACCGTGGCAGGCTCG 4260  
 DB 4201 ATTAATGCTGCTTAATGTGACATTAATACAGGCACTTTAAACCAACCGTGGCAGGCTCG 4260  
 QY 4261 GATATTAAGCAACCAAGCGGCACTTTGGTATTATTAACGCAAAAGATGCTAAAGTAAATGTT 4320  
 DB 4261 GATATTAAGCAACCAAGCGGCACTTTGGTATTATTAACGCAAAAGATGCTAAAGTAAATGTT 4320  
 QY 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAAGCAAGGCGCTCTGCTAGTGTG 4380  
 DB 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAAGCAAGGCGCTCTGCTAGTGTG 4380  
 QY 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCATCTGGGATTTAAACACAGTAAATGGGTTA 4440  
 DB 4381 ACTGCGGCAACCTCAAGCAGTGTGAATATCATCTGGGATTTAAACACAGTAAATGGGTTA 4440  
 QY 4441 AATATCATTTTCGAAAGATGGTAGAAGACTGTGAGCAAGTAAATTTGAAGCAAGGCGCTCT 4500  
 DB 4441 AATATCATTTTCGAAAGATGGTAGAAGACTGTGAGCAAGTAAATTTGAAGCAAGGCGCTCT 4500  
 QY 4501 AATATATATCAGCAGGCTGTAGCAAGTGTAGAAGAGTAAATTTGAAGCAAGGCGCTCT 4560  
 DB 4501 AATATATATCAGCAGGCTGTAGCAAGTGTAGAAGAGTAAATTTGAAGCAAGGCGCTCT 4560  
 QY 4561 GAAAGTAAAGATTTTATCTGATGAAGAAAGAGAAACATTTAGCTAACTTTGGTGAAGT 4620  
 DB 4561 GAAAGTAAAGATTTTATCTGATGAAGAAAGAGAAACATTTAGCTAACTTTGGTGAAGT 4620  
 QY 4621 GCTGTACGTTTGTGAGCCAAATAATTAACAGTCAATACACAAAATGAATTTTACA 4680  
 DB 4621 GCTGTACGTTTGTGAGCCAAATAATTAACAGTCAATACACAAAATGAATTTTACA 4680  
 QY 4681 ACCAGCGCTCAAGTCAAGTGAATTTTCTGAAGTGAAGCGGTGTTTCTCAAGTGGTAAAT 4740  
 DB 4681 ACCAGCGCTCAAGTCAAGTGAATTTTCTGAAGTGAAGCGGTGTTTCTCAAGTGGTAAAT 4740  
 QY 4741 GCGCAGCAGTATGTACCAATTTGCTGACAGTGAAGCGGCTAGTCAAGTAAATTTGCAAG 4800

```

Db 4741 GCGCAGGATGATGACCAATGTTGCTGACGATGGACCGCTAGTCAGTAATTGACAAG 4800
QY 4801 GTAGATTTCAATCCGCAATGAAGTCAATTTTTCGCTATTTTACTGTTGGGTTAA 4860
Db 4801 GTAGATTTCAATCCGCAATGAAGTCAATTTTTCGCTATTTTACTGTTGGGTTAA 4860
QY 4861 GTTCAAGTACGGGCTTTACCCATCTGTGTAATAAATACGAGAGTAACAATAAGTATTTT 4920
Db 4861 GTTCAAGTACGGGCTTTACCCATCTGTGTAATAAATACGAGAGTAACAATAAGTATTTT 4920
QY 4921 AACAGGTTATTTATG 4937
Db 4921 AACAGGTTATTTATG 4937

```

## RESULT 8

```

US-09-942-70
; Sequence 70, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206.942
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167.568
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
; LENGTH: 4937
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-09-206-942-70

```

```

Query Match      100.0%; Score 4937; DB 4; Length 4937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATATACAGATTAATAAATAAATCAAGATTTTGTGATGACAAACAAATACAA 60
Db 1 TAAATATACAGATTAATAAATAAATCAAGATTTTGTGATGACAAACAAATACAA 60
QY 61 CACCTTTTTCGAGTCTATATGCAATATTTTAAAAAATAGTATAAATCGGCATATA 120
Db 61 CACCTTTTTCGAGTCTATATGCAATATTTTAAAAAATAGTATAAATCGGCATATA 120
QY 121 AATGATATAATCTTTCATCTTTTCATCTTTTATCTTTCATCTTTTCATCTTTTCAT 180
Db 121 AATGATATAATCTTTCATCTTTTCATCTTTTATCTTTCATCTTTTCATCTTTTCAT 180
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 240
Db 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCT 240
QY 241 CACATGAATATGATGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 241 CACATGAATATGATGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
QY 301 GAGCGCAATATGATGAAGTAATTTAATGTTTCAACTAACCTTTAGGAGAAATATGAACA 360
Db 301 GAGCGCAATATGATGAAGTAATTTAATGTTTCAACTAACCTTTAGGAGAAATATGAACA 360
QY 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGGTGTGCTGTCTGCTGAATGGCA 420
Db 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGGTGTGCTGTCTGCTGAATGGCA 420
QY 421 CGGGGTTGTGACCATTTCCACAGAAAAGGCTTCGCTATGTTACTATCTTTAGGTGTAAC 480

```

```

Db 421 CGGGGTTGTGACCATTTCCACAGAAAAAGGCTTCGCTATGTTACTATCTTTAGGTGTAAC 480
QY 481 CACTTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540
Db 481 CACTTTAGCGTTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAACATCTATTCCA 540
QY 541 CAATCTCTTTTACGACGCGGCTTACAGGAATGGATGTAGTACACGCGCAGCCACTATG 600
Db 541 CAATCTCTTTTACGACGCGGCTTACAGGAATGGATGTAGTACACGCGCAGCCACTATG 600
QY 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGAGCTATCAATTAATTTGGAAA 660
Db 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGAGCTATCAATTAATTTGGAAA 660
QY 661 CAATTTAATCATGACCAAAATGAAATGGTGGCAGTGTGAGCTATCAATTAATTTGGAAA 720
Db 661 CAATTTAATCATGACCAAAATGAAATGGTGGCAGTGTGAGCTATCAATTAATTTGGAAA 720
QY 721 GTATTCAACCGTGTGTTACATCTTACCAAAATCTCCCAATTTAAAGGGATTTTAGATCTAAC 780
Db 721 GTATTCAACCGTGTGTTACATCTTACCAAAATCTCCCAATTTAAAGGGATTTTAGATCTAAC 780
QY 781 GGACAAAGTCTTTTAAATCAACCCAAATGCTATCACAATAGGTAAGAGCGCAATTTAAAC 840
Db 781 GGACAAAGTCTTTTAAATCAACCCAAATGCTATCACAATAGGTAAGAGCGCAATTTAAAC 840
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900
Db 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900
QY 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGCAAAATGCTGAATTCACGCTTAAAT 960
Db 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGCAAAATGCTGAATTCACGCTTAAAT 960
QY 961 ACTGTCGTTAAAGACGCGAGTGTAAATCTTATGTTGGCAAAAGTGAAGAGGAGGTTG 1020
Db 961 ACTGTCGTTAAAGACGCGAGTGTAAATCTTATGTTGGCAAAAGTGAAGAGGAGGTTG 1020
QY 1021 ATTACGCTAAATGTTGCGAGCATTTCTTACTGCGAGGCAAAATACCATTCAGCGAT 1080
Db 1021 ATTACGCTAAATGTTGCGAGCATTTCTTACTGCGAGGCAAAATACCATTCAGCGAT 1080
QY 1081 ATAATAAACCCACCACTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGCTCAATCTG 1140
Db 1081 ATAATAAACCCACCACTTACTTACAGCATTTGCGCGCTGAAATGAAGCGGCTCAATCTG 1140
QY 1141 GCGCATATTTTCCCAAGGCGGTAAACATTAATGTCGCTGTCACATTTTCGAAACCAA 1200
Db 1141 GCGCATATTTTCCCAAGGCGGTAAACATTAATGTCGCTGTCACATTTTCGAAACCAA 1200
QY 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATTAAGGGGCAATATTTGTTTCCGCC 1260
Db 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATTAAGGGGCAATATTTGTTTCCGCC 1260
QY 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320
Db 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGC 1320
QY 1321 GGCAAGCTGATGATTTACAGCGGATTAAGTCAATTAAGAGAGGTGAGTATTCGACCTT 1380
Db 1321 GGCAAGCTGATGATTTACAGCGGATTAAGTCAATTAAGAGAGGTGAGTATTCGACCTT 1380
QY 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCGCAAGGTAAGAAC 1440
Db 1381 TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGCGCAAGGTAAGAAC 1440
QY 1441 GGCATTCATTTAGCAAAAGCAACCTCTTTAGAAAAGGCTCAACCATCAATGATTCAGGC 1500
Db 1441 GGCATTCATTTAGCAAAAGCAACCTCTTTAGAAAAGGCTCAACCATCAATGATTCAGGC 1500
QY 1501 AAAGAAAAGGCGGACGCGCTATTTGTTGGGCGGATATTCGCTTAATTCAGCGCAATATT 1560
Db 1501 AAAGAAAAGGCGGACGCGCTATTTGTTGGGCGGATATTCGCTTAATTCAGCGCAATATT 1560

```







```
Db 3721 ATTACTGCAAAAATGTAGAAATGTAACAAAGATATTACTTCTCTCAAAACAGTAATATC 3780
Qy 3781 ACCGGCTGGAAAGGTTACCAACACAGAGCTGACACATTAACGCAACAAATGCGCAA 3840
Db 3781 ACCGGCTGGAAAGGTTACCAACACAGAGCTGACACATTAACGCAACAAATGCGCAA 3840
Qy 3841 GCAAGTATTACAAACCAACAGAGTGATATCAGCGGTACGATTTCCGGGTAAACAGGTAAGT 3900
Db 3841 GCAAGTATTACAAACCAACAGAGTGATATCAGCGGTACGATTTCCGGGTAAACAGGTAAGT 3900
Qy 3901 GTTACGCGACTGGTATTAACACATTAATTCGGCTCAAAATTTGAAGCGAAATCGGCT 3960
Db 3901 GTTACGCGACTGGTATTAACACATTAATTCGGCTCAAAATTTGAAGCGAAATCGGCT 3960
Qy 3961 GAGGCTAATGTACAAAGTCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020
Db 3961 GAGGCTAATGTACAAAGTCAACAGGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020
Qy 4021 AATGTTACCGCAACCGCTGGGATTTAACAGTTTGGCAATGGCGGAGAAATTAATCGGACA 4080
Db 4021 AATGTTACCGCAACCGCTGGGATTTAACAGTTTGGCAATGGCGGAGAAATTAATCGGACA 4080
Qy 4081 GAAGAGCTGCAACCTTAACCGCAACAGGTAATTTGACTACTGAAGCGGTTCTAGC 4140
Db 4081 GAAGAGCTGCAACCTTAACCGCAACAGGTAATTTGACTACTGAAGCGGTTCTAGC 4140
Qy 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGATGCTAGCAGGCAAGC 4200
Db 4141 ATCACTTCAACTAAGGCTCAGGTAGACCTCTTGGCTCAGATGCTAGCAGGCAAGC 4200
Qy 4201 ATTAATGCTGTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Db 4201 ATTAATGCTGTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4260
Qy 4261 GATATTAAGCAACCGGCAACCTTGGTATTAAACGCAAGGTAAGTAAATGCT 4320
Db 4261 GATATTAAGCAACCGGCAACCTTGGTATTAAACGCAAGGTAAGTAAATGCT 4320
Qy 4321 GATCATCAGGTAGTATGACAGAGTGAATGAGTCAAGCAAGCGGCTCTGGTAGTGTG 4380
Db 4321 GATCATCAGGTAGTATGACAGAGTGAATGAGTCAAGCAAGCGGCTCTGGTAGTGTG 4380
Qy 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGGTTA 4440
Db 4381 ACTCGGCAACCTCAAGCAGTGTGAATATCACTGGGATTTAAACACAGTAAATGGGTTA 4440
Qy 4441 AATATCATTTGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATGAGGTG 4500
Db 4441 AATATCATTTGAAAGATGTTAGAAACACTGTGCGCTTAAGAGGCAAGGAAATGAGGTG 4500
Qy 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTGAATTTGAAGCGAAACGGTCTT 4560
Db 4501 AATATATCCAGCCAGGTGTAGCAAGTGTAGAAAGTGAATTTGAAGCGAAACGGTCTT 4560
Qy 4561 GAAAGATGAAAGATTTATCTGATGAGAAAGAGAAACATTAAGTAACTGTTGTTAAGT 4620
Db 4561 GAAAGATGAAAGATTTATCTGATGAGAAAGAGAAACATTAAGTAACTGTTGTTAAGT 4620
Qy 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTAAGTCAATACAAATTAAGTCAAT 4680
Db 4621 GCTGTACGTTTGTGAGCCAAATATACAAATTAAGTCAATACAAATTAAGTCAAT 4680
Qy 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTCAAGGTAAAGCGTGTCTCAAGTGAAT 4740
Db 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTCAAGGTAAAGCGTGTCTCAAGTGAAT 4740
Qy 4741 GCGCACGAGTATGACCAATGCTGACGATGGACGCGTAGTCAAGTAAATGACAAG 4800
Db 4741 GCGCACGAGTATGACCAATGCTGACGATGGACGCGTAGTCAAGTAAATGACAAG 4800
Qy 4801 GTAGATTTTCATCTGCAAGTGAATTTATTTTCTGATTTATTTTCTGTTGTTGTTGTTGTT 4860
Db 4801 GTAGATTTTCATCTGCAAGTGAATTTATTTTCTGATTTATTTTCTGTTGTTGTTGTTGTT
```

```
Db 4801 GTAGATTTTCATCTGCAAGTGAATTTATTTTCTGATTTATTTTCTGTTGTTGTTGTT 4860
Qy 4861 GTTCAAGTACGGCTTTTACCATCTTGTAAAAAATTAACGAGAAATCAATCAATCAAT 4920
Db 4861 GTTCAAGTACGGCTTTTACCATCTTGTAAAAAATTAACGAGAAATCAATCAATCAAT 4920
Qy 4921 AACAGGTTATTTATTTATG 4937
Db 4921 AACAGGTTATTTATTTATG 4937

RESULT 9
US-08-038-682-6
; Sequence 6, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-6

Query Match 97.6%; Score 4820.2; DB 1; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 1 TAAATATACAGATAATAAATAAATCAAGATTTTGTGATGACAAACAACAAATACAA 60
Db 441 TAAATATACAGATAATAAATAAATAAATCAAGATTTTGTGATGACAAACAACAAATACAA 500
Qy 61 CACCTTTTTCGAGCTCTATATCAATATTTTAAAAAATAGTAAATACCGCATATAA 120
Db 501 CACCTTTTTCGAGCTCTATATCAATATTTTAAAAAATAGTAAATACCGCATATAA 560
Qy 121 AATGATATATCTTTCATCTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAAT 180
Db 561 AATGATATATCTTTCATCTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAAT 620
Qy 181 CTTTCATCTTTCATCTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCTTTT 240
Db 621 CTTTCATCTTTCATCTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCTTT 680
```





QY 4621 GCTGTACGTTTGTGGAGCAATATACAAATTACAGTCAATACACAAATGAATTTACA 4680  
Db 5061 GCTGTACGTTTGTGGAGCAATATACAAATTACAGTCAATACACAAATGAATTTACA 5120  
QY 4681 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGGCGTGTCTCAAGTGGTAAT 4740  
Db 5121 ACCAGACCGTCAAGTCAAGTGAATTTCTGAAGTAAAGGCGTGTCTCAAGTGGTAAT 5180  
QY 4741 GCGCAGCAGTATGTACCAATGCTGACGATGACAGCCGTAGTCAAGTAAATGACAAG 4800  
Db 5181 GCGCAGCAGTATGTACCAATGCTGACGATGACAGCCGTAGTCAAGTAAATGACAAG 5240  
QY 4801 GTAGATTTCATCTGCAATGAAGTCAATTTATTTTCGTATTTATTTTCGTGTTGTTAAA 4860  
Db 5241 GTAGATTTCATCTGCAATGAAGTCAATTTATTTTCGTATTTATTTTCGTGTTGTTAAA 5300  
QY 4861 GTTCAGTACGGCTTTACCCATCTGTAAGAAATACGAGAAATACAAATAAGTATTTT 4920  
Db 5301 GTTCAGTACGGCTTTACCCATCTGTAAGAAATACGAGAAATACAAATAAGTATTTT 5360  
QY 4921 AACAGGTTATTTATG 4937  
Db 5361 AACAGGTTATTTATG 5377

## RESULT 10

US-08-302-832-6  
; Sequence 6, Application US/08302832  
; Patent No. 5603938  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/302,832  
; FILING DATE: 16-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US pct/us93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstesser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-404  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9323 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-302-832-6

Query Match 97.6%; Score 4820.2; DB 1; Length 9323;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 1 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAATATACAA 60  
Db 441 TAAATATACAGATATAAATAAATAAATCAAGATTTTGTGATGACAAACAATATACAA 500  
QY 61 CACCTTTTTCAGCTATATATGCAAAATATTTTAAAAAATAGTATAATCGCCATATAA 120  
Db 501 CACCTTTTTCAGCTATATATGCAAAATATTTTAAAAAATAGTATAATCGCCATATAA 560  
QY 121 AATGGTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180  
Db 561 AATGGTAAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620  
QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680  
QY 241 CACATGAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 681 CACATGAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740  
QY 301 GAAAGCAATGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 360  
Db 741 GAAAGCAATGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 800  
QY 361 ATATATGCTCTCAAAATTCAGCAACGCTGATGCTTTGGTGTGCTGTCTGCTGAATGGCA 420  
Db 801 ATATATGCTCTCAAAATTCAGCAACGCTGATGCTTTGGTGTGCTGTCTGCTGAATGGCA 860  
QY 421 CGGGGTGTGACCATTCACAGAAAAGGCTCCCGCTATGTTACTATCTTTAGGTGTAAAC 480  
Db 861 CGGGGTGTGACCATTCACAGAAAAGGCTCCCGCTATGTTACTATCTTTAGGTGTAAAC 920  
QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAAACCTATTTTAC 540  
Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGGTGTAAACCTATTTTAC 980  
QY 541 CAATCTGTTTATGCAAGCGGCTTACAAGGAATGGATGATACAGCGGACACGCCACTATG 600  
Db 981 CAATCTGTTTATGCAAGCGGCTTACAAGGAATGGATGATACAGCGGACACGCCACTATG 1040  
QY 601 CAAGTAGATGGTAAATAAACCCTATATCCGCAACAGTGTGAGCGCTATCTAATTTGGA 660  
Db 1041 AAGAAAACAAGTAATAAACCCTATATCCGCAACAGTGTGAGCGCTATCTAATTTGGA 1100  
QY 661 CAATTTAACATCGACCAAAATGAATGGTGCAGTGTGTTTACAAGAAAACAACACTCCGC 720  
Db 1101 CAATTTAACATCGACCAAAATGAATGGTGCAGTGTGTTTACAAGAAAACAACACTCCGC 1160  
QY 721 GTATTCACCGTGTACATCTACCAAAATCTCCCAATTAAGAGGATTTTATAGATTTTAA 780  
Db 1161 GTATTCACCGTGTACATCTACCAAAATCTCCCAATTAAGAGGATTTTATAGATTTTAA 1220  
QY 781 GGACAGTCTTTTATACACCAAAATGGTATCACAATAGTAAAGCAATTTATTAAC 840  
Db 1221 GGACAGTCTTTTATACACCAAAATGGTATCACAATAGTAAAGCAATTTATTAAC 1280  
QY 841 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 900  
Db 1281 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGCGCGTAAT 1340  
QY 901 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATCATCGGTTAAT 960  
Db 1341 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATCATCGGTTAAT 1400  
QY 961 ACTGTCGTAAGACGCGCAGTGTAAATCTTATTTGGTGGCAAGTGAAGAAACGAGGTTG 1020  
Db 1401 ACTGTCGTAAGACGCGCAGTGTAAATCTTATTTGGTGGCAAGTGAAGAAACGAGGTTG 1460

QY 1021 ATTACGGTAATGGTGGCAGCATTTCTTTTACTCGCAGGCGCAAAAATCACCATCAGCGAT 1080  
DB 1461 ATTACGGTAATGGTGGCAGCATTTCTTTACTCGCAGGCGCAAAAATCACCATCAGCGAT 1520  
QY 1081 ATAATTAACCCCAACCAATTAATTAACAGCATTCGCGCCCTGAAATGAAGCGGTCAATCTG 1140  
DB 1521 ATAATTAACCCCAACCAATTAATTAACAGCATTCGCGCCCTGAAATGAAGCGGTCAATCTG 1580  
QY 1141 GCGCATATTTTCCCAAGGCGGTAAACATTAATGTCCGTGCTGCCACTATTCGAACCA 1200  
DB 1581 GCGCATATTTTCCCAAGGCGGTAAACATTAATGTCCGTGCTGCCACTATTCGAACCA 1640  
QY 1201 GGTAACTTTCTGCTGATTTCTGTAAGCAAGATTAAGCGCGCAATATTTCTTTCCGCC 1260  
DB 1641 GGTAACTTTCTGCTGATTTCTGTAAGCAAGATTAAGCGCGCAATATTTCTTTCCGCC 1700  
QY 1261 AAAGAGGGTGAAGCGCAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320  
DB 1701 AAAGAGGGTGAAGCGCAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1760  
QY 1321 GCAAGCTGATGATTAACAGCGATTAAGTCAATTAAGCAAGTGCAGTATACGACCTT 1380  
DB 1761 GCAAGCTGATGATTAAGTCAATTAAGTCAATTAAGCAAGTGCAGTATACGACCTT 1820  
QY 1381 TCAGCTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGCGAAGGTAAAAAC 1440  
DB 1821 TCAGCTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGCGAAGGTAAAAAC 1880  
QY 1441 GCAATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500  
DB 1881 GCAATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1940  
QY 1501 AAAGAAAAAGGCGGACGGCTATTTGGTGGGCGATATTCGCTTAATTTAGCGCAATATTT 1560  
DB 1941 AAAGAAAAAGGCGGACGGCTATTTGGTGGGCGATATTCGCTTAATTTAGCGCAATATTT 2000  
QY 1561 AACGCTCAAGGTAGTGTGATTCGCTAAAACCGGTGTTTGTGAGACATCGGGGCAT 1620  
DB 2001 AACGCTCAAGGTAGTGTGATTCGCTAAAACCGGTGTTTGTGAGACATCGGGGCAT 2060  
QY 1621 TATTTATCCATTTGACGCAATGCAATTTGTTAAAACAAAGAGTGGTCTAGACCTGAT 2120  
DB 2061 TATTTATCCATTTGACGCAATGCAATTTGTTAAAACAAAGAGTGGTCTAGACCTGAT 2180  
QY 1681 GATGTAAACATTTGAAGCGGAGACCCCTTCCGCAATTAATACCGGTATTAATGATGAATTC 1740  
DB 2121 GATGTAAACATTTGAAGCGGAGACCCCTTCCGCAATTAATACCGGTATTAATGATGAATTC 2180  
QY 1741 CCACAGGCGCGTGAAGCAAGCGACCTTAATAAATAATAGCAACTCAAAACCAAGCCTA 1800  
DB 2181 CCACAGGCGCGTGAAGCAAGCGACCTTAATAAATAATAGCAACTCAAAACCAAGCCTA 2240  
QY 1801 ACCAATACCACTATTTCAAAATTAATCTGAAAACGCTTGGACCAATTAATTAAGCGGATCA 1860  
DB 2241 ACCAATACCACTATTTCAAAATTAATCTGAAAACGCTTGGACCAATTAATTAAGCGGATCA 2300  
QY 1861 AGAAGACTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTTCTCCAT 1920  
DB 2301 AGAAGACTTACCGTTAATAGCTCAATCAACATCGGAAGCACTCCCACTTAATTTCTCCAT 2360  
QY 1921 AGTAAGCTCAGCGTGGCGGAGGCTTACAGATTGATGGAGATATTAATTTCTAAAGCGGA 1980  
DB 2361 AGTAAGCTCAGCGTGGCGGAGGCTTACAGATTGATGGAGATATTAATTTCTAAAGCGGA 2420  
QY 1981 AATTTAAACCAATTTATTTCTGGCGGATGGTGTGATGTCATATAAATATTAACGTTGATCAG 2040  
DB 2421 AATTTAAACCAATTTATTTCTGGCGGATGGTGTGATGTCATATAAATATTAACGTTGATCAG 2480  
QY 2041 GGTTTTTTAATTAATTAACCGCGCTTCGCTAGCTTTTGAAGGTGGAAATTAACAAACGCGC 2100  
DB 2481 GGTTTTTTAATTAATTAACCGCGCTTCGCTAGCTTTTGAAGGTGGAAATTAACAAACGCGC 2540  
QY 2101 GACGCGCAATGCTAAAAATTTGTCGCCCGGACGCTGTACCATTTACAGAGGAGGAAAA 2160

DB 2541 GACGCGCAATGCTAAAAATTTGTCGCCAGGCGACTGTAAACATTTACAGGAGGAGAAAA 2600  
QY 2161 GATTTTCAGGCTAAACAGGTATCTTTAAACGCAAGCGGTAAAGTCTGTAATATCATTTCA 2220  
DB 2601 GATTTTCAGGCTAAACAGGTATCTTTAAACGCAAGCGGTAAAGTCTGTAATATCATTTCA 2660  
QY 2221 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGCACAATTAACATATCTGGGAATAACA 2280  
DB 2661 TCAGTGAATTAATTTAAACCCCAATCTTTAGTGGCACAATTAACATATCTGGGAATAACA 2720  
QY 2281 ATTAAACCAACTACGAGAAAGACACCTCGTATTCGCAAAACGACCATGATTCGCACTGG 2340  
DB 2721 ATTAAACCAACTACGAGAAAGACACCTCGTATTCGCAAAACGACCATGATTCGCACTGG 2780  
QY 2341 AACGTCAGTCTCTTTAATCTAGAGACAGGCGCAATTTTACCTTTAATTAATACATTTCA 2400  
DB 2781 AACGTCAGTCTCTTTAATCTAGAGACAGGCGCAATTTTACCTTTAATTAATACATTTCA 2840  
QY 2401 AGCAATAGCAAAAGGCTTAACAAACACAGTATAGAAGCTCTGAGGGGTGAATTTAAACGCC 2460  
DB 2841 AGCAATAGCAAAAGGCTTAACAAACACAGTATAGAAGCTCTGAGGGGTGAATTTAAACGCC 2900  
QY 2461 GTAAATGGCAACATGTCATTTCAATCTCAAGAGAGGCGAAAGTTAATTTCAATTTAAAA 2520  
DB 2901 GTAAATGGCAACATGTCATTTCAATCTCAAGAGAGGCGAAAGTTAATTTCAATTTAAAA 2960  
QY 2521 CCAAGGAGACATGAACACACAGCAAACTTTTACCAATTCGGTCTTTAGCCCAATATCACA 2580  
DB 2961 CCAAGGAGACATGAACACACAGCAAACTTTTACCAATTCGGTCTTTAGCCCAATATCACA 3020  
QY 2581 GCCACTTGGTGGGCGCTCTGTTTTTGTATATATATGCCAACCATTTTGGCAGAGGGCT 2640  
DB 3021 GCCACTTGGTGGGCGCTCTGTTTTTGTATATATATGCCAACCATTTTGGCAGAGGGCT 3080  
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTCTAAGCGGCTAATTTTACCTTTAAATTTCCCAT 2700  
DB 3081 GAGTTAAAAATGAGTGAATTAATATCTCTAAGCGGCTAATTTTACCTTTAAATTTCCCAT 3140  
QY 2701 GTTCGCGCGATGACGCTTTTAAATCAACAAAGACTTAAACATTAATTAACCAATTTCA 2760  
DB 3141 GTTCGCGCGATGACGCTTTTAAATCAACAAAGACTTAAACATTAATTAACCAATTTCA 3200  
QY 2761 AATTTACGCTCAGACAGAGAAAGATGATTTTATGACGGGTAGCGCAAGTGCATC 2820  
DB 3201 AATTTACGCTCAGACAGAGAAAGATGATTTTATGACGGGTAGCGCAAGTGCATC 3260  
QY 2821 AATTTACGCTCAGACATATCCATTTCTGGCGGTAAATGTACCCCTTGGTGGACAAACTCA 2880  
DB 3261 AATTTACGCTCAGACATATCCATTTCTGGCGGTAAATGTACCCCTTGGTGGACAAACTCA 3320  
QY 2881 AGCAGCAGCTTACGGGGAATTAATCTATCGAAGAGCAGCAATGTTAGCCTAGAGCC 2940  
DB 3321 AGCAGCAGCTTACGGGGAATTAATCTATCGAAGAGCAGCAATGTTAGCCTAGAGCC 3380  
QY 2941 AATTAACGCGCTAATCAGCAAAACATTAAGGATAGAGTTATAAACTTGGCAGTTGCTC 3000  
DB 3381 AATTAACGCGCTAATCAGCAAAACATTAAGGATAGAGTTATAAACTTGGCAGTTGCTC 3440  
QY 3001 GTTAATGGGAGTTTAAAGTTTAACTGGGGAATTAATTAAGGCAATCTCACTATT 3060  
DB 3441 GTTAATGGGAGTTTAAAGTTTAACTGGGGAATTAATTAAGGCAATCTCACTATT 3500  
QY 3061 TCAGAAAGCGGCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3120  
DB 3501 TCAGAAAGCGGCACTTTTAAAGGAAAGACTAGAGATACCTTAATATACCGGCAATTTT 3560  
QY 3121 ACCAATTAAGCAGTGGCGAAATTAATATACCAAGAGTGGTAAACCTTGGCAATGTT 3180  
DB 3561 ACCAATTAAGCAGTGGCGAAATTAATATACCAAGAGTGGTAAACCTTGGCAATGTT 3620  
QY 3181 ACCAATGATGTTGATTTTAAACATTTACCTCAGCTAAACGCAACCAAGAGATCATC 3240



3621	DB	ACCAATGATGGTGTAAATTAACATTTACCTCAGCTTAACGCAACCAAGAAGCATCATC	3680
3241	QY	GGCGGAGATATATCAACAACAAAAAGGAAGCTTAATATTTACAGACAGTAAATAATGATGCT	3300
3681	DB	GGCGGAGATATATCAACAACAAAAAGGAAGCTTAATATTTACAGACAGTAAATAATGATGCT	3740
3301	QY	GAATTCCAAAATTTGGCGGCAATATCTCGCAAAAGAAGGCAACCTCAGATTTCTTCGAT	3360
3741	DB	GAATTCCAAAATTTGGCGGCAATATCTCGCAAAAGAAGGCAACCTCAGATTTCTTCGAT	3800
3361	QY	AAAAATTAATATCACCAACACAGATAACAATCAAAAAGGGTATTGATGGAGAGAGACTTAGT	3420
3801	DB	AAAAATTAATATCACCAACACAGATAACAATCAAAAAGGGTATTGATGGAGAGAGACTTAGT	3860
3421	QY	TCAGATGCGACAGTAATGCGCAACCTAACTATTATAACCAAAAGAAATTTGAAATTTGACAGAA	3480
3861	DB	TCAGATGCGACAGTAATGCGCAACCTAACTATTATAACCAAAAGAAATTTGAAATTTGACAGAA	3920
3481	QY	GACCTTAAGTATTTTCAGGCTTTCAATTAAGCAGAGATTTACAGCCAAAGATGTTAGAGATTTA	3540
3921	DB	GACCTTAAGTATTTTCAGGCTTTCAATTAAGCAGAGATTTACAGCCAAAGATGTTAGAGATTTA	3980
3541	QY	ACTATTGGCAACAGTAATGACGGTTAACAGCGGTGCCGAAGCCAAACACAGTAACTTTTAAC	3600
3981	DB	ACTATTGGCAACAGTAATGACGGTTAACAGCGGTGCCGAAGCCAAACACAGTAACTTTTAAC	4040
3601	QY	ATGTTTAAAGATTCAAAATCTCGCTGACGGTCAACATGTGCACACTAAATAGCAAGTG	3660
4041	DB	ATGTTTAAAGATTCAAAATCTCGCTGACGGTCAACATGTGCACACTAAATAGCAAGTG	4100
3661	QY	AAACATCTACAGCAATGGCGGACGTGAAGCAANTAGCGCAACGATACCGGCTTAATC	3720
4101	DB	AAACATCTACAGCAATGGCGGACGTGAAGCAANTAGCGCAACGATACCGGCTTAATC	4160
3721	QY	ATTACTGCAAAAATGTAGAAGTAACAAGATATTACTTCTCTCAAAACAGTAAATATC	3780
4161	DB	ATTACTGCAAAAATGTAGAAGTAACAAGATATTACTTCTCTCAAAACAGTAAATATC	4220
3781	QY	ACCGGCTCGGAAAAAGTTTACCACACAGCAGGCTCGACCAATTAACGCAACAATGGCAA	3840
4221	DB	ACCGGCTCGGAAAAAGTTTACCACACAGCAGGCTCGACCAATTAACGCAACAATGGCAA	4280
3841	QY	GCAAGTATTACAACCAACACAGGTGATACAGCGGTACGATTTCCGGTAAACCGTAAAT	3900
4281	DB	GCAAGTATTACAACCAACACAGGTGATACAGCGGTACGATTTCCGGTAAACCGTAAAT	4340
3901	QY	GTTTAGCGGACTGGTGATTTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGT	3960
4341	DB	GTTTAGCGGACTGGTGATTTAACCACTAAATCCGGCTCAAAAATTTGAAGCGAAATCGGT	4400
3961	QY	GAGGCTAATGTAAACAAAGTGAACAGGTACAATTTGGCGGTACAATTTCCGGTAAACGGT	4020
4401	DB	GAGGCTAATGTAAACAAAGTGAACAGGTACAATTTGGCGGTACAATTTCCGGTAAACGGT	4460
4021	QY	AATGTTACGGCAACCGCTGGCGATTTTAACAGTTGGGAATGCGCAGAAATTAATCGACA	4080
4461	DB	AATGTTACGGCAACCGCTGGCGATTTTAACAGTTGGGAATGCGCAGAAATTAATCGACA	4520
4081	QY	GAAGGAGCTGAACCTTTAACCGCACAGGGAATACCTTGACTACTGAACCGGTTCTAGC	4140
4521	DB	GAAGGAGCTGAACCTTTAACCGCACAGGGAATACCTTGACTACTGAACCGGTTCTAGC	4580
4141	QY	ATCATTCAACTAAGGGTCAAGGTAGACCTCTTTGGCTCAGAAATGGTATCGACAGAGAC	4200
4581	DB	ATCATTCAACTAAGGGTCAAGGTAGACCTCTTTGGCTCAGAAATGGTATCGACAGAGAC	4640
4201	QY	ATTAATGCTGCTTAATGTGACATTTAAATACTACAGGCACCTTTAACCCAGGCTCG	4260
4641	DB	ATTAATGCTGCTTAATGTGACATTTAAATACTACAGGCACCTTTAACCCAGGCTCG	4700
4261	QY	GATATTAAAGCAACACAGGGCACCTTGTTTATTAAACCAAAAGATGCTAAGCTTAAATGCT	4320
4701	DB	GATATTAAAGCAACACAGGGCACCTTGTTTATTAAACCAAAAGATGCTAAGCTTAAATGCT	4760

## RESULT 11

RESULT II  
US-08-530-198-6

US-08-530-198-6  
: Sequence 6, Application us/08530198

: Sequence 6, Application  
: Patent No. 5869065

Patent No. 5869063

GENERAL INFORMATION:

;; GENERAL INFORMATION:  
: APPLICANT: BARENKAMP, STEPHEN J

APPLICANT: BARENKAMP, STEPHEN J  
APPLICANT: ST. GEME III. JOSEPH W

APPLICANT: ST. GEME III, JOSEPH W  
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

;; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFAC

; TITLE OF INVENTION: OF  
 . NUMBER OF SEQUENCES: 8

; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:

; CORRESPONDENCE ADDRESS: shoemaker and Mattare. Ltd

ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: 2001 Jefferson Davis Hwy. 1203 Crystal Plaza

STREET: 2001 JE  
STREET: 1

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS

; SOFTWARE: PatentIn Release #1.0, V

; CURRENT APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:  
 NAME: BERKSTRESSER, JERRY W  
 REGISTRATION NUMBER: 22,651  
 REFERENCE/DOCKET NUMBER: JWB-1186  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 415-0813  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9323 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-530-198-6

Query Match 97.6%; Score 4820.2; DB 2; Length 9323;  
 Best Local Similarity 98.5%; Pred. No. 0;  
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGATATATAAATAAATCAAGATTTTGTGATGACAAACAAATTCACAA 60  
 DB 441 TAAATATACAGATATATAAATAAATCAAGATTTTGTGATGACAAACAAATTCACAA 500

QY 61 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATAA 120  
 DB 501 CACCTTTTTCAGCTCTATATGCAAAATATTTTAAAAAATAGTATATAATCCGCCATATAA 560

QY 121 AATGGTATATCTTTTCATCTTTTCAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 180  
 DB 561 AATGGTATATCTTTTCATCTTTTCAATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620

QY 181 CTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 240  
 DB 621 CTTCATCTTTTCATCTTTTCAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680

QY 241 CACATGAAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300  
 DB 681 CACATGAAATGATGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740

QY 301 GACGCAATGATGAATTAATTTTCACTTAACCTTTAGGAGAAATATGACAAAG 360  
 DB 741 GACGCAATGATGAATTAATTTTCACTTAACCTTTAGGAGAAATATGACAAAG 800

QY 361 ATATATCGTCTCAAAATTCAGCAACGCTGAATCTTTGGTGTGCTGTCTGAATTTGGCA 420  
 DB 801 ATATATCGTCTCAAAATTCAGCAACGCTGAATCTTTGGTGTGCTGTCTGAATTTGGCA 860

QY 421 CGGGTGTGACCATTCACAGAAAGGCTTCGCGTATGTTACTATCTTTTATGTTGTTAATAC 480  
 DB 861 CGGGTGTGACCATTCACAGAAAGGCTTCGCGTATGTTACTATCTTTTATGTTGTTAATAC 920

QY 481 CACTTAGCGTTAAAGCCACCTTTCCGCTATGTTACTATCTTTAGTGTAACTATCTATTCCA 540  
 DB 921 CACTTAGCGTTAAAGCCACCTTTCCGCTATGTTACTATCTTTAGTGTAACTATCTATTCCA 980

QY 541 CAATCTGTTTACGACGGCTTTACAGGAATGATGTAGTACACGGCACAGCCACTATG 600  
 DB 981 CAATCTGTTTACGACGGCTTTACAGGAATGATGTAGTACACGGCACAGCCACTATG 1040

QY 601 CAAGTATAGTGTATATAAACCCTTTATCCGCAACAGTGTGACGCTATCATTAATTCGAAA 660  
 DB 1041 AAGAAAACAAGTATATAAACCCTTTATCCGCAACAGTGTGACGCTATCATTAATTCGAAA 1100

QY 661 CAAATTAACATCGACCAAAATGAATGGTGTGAGTGTATCAAGAAAACAACTCCGCC 720  
 DB 1101 CAAATTAACATCGACCAAAATGAATGGTGTGAGTGTATCAAGAAAACAACTCCGCC 1160

QY 721 GTATTCACCGGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTAGATCTCTAC 780  
 DB 1161 GTATTCACCGGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTAGATCTCTAC 1220

QY 781 GGNACAGCTTTTATCAACCAAAATGGTATCACATATAGGTAAGGACGCAATTTATTAAC 840

DB 1221 GGCAAGTCTTTTAAATCAACCCAAATGGTATCACAATAGTAAAGACGCAATTTATTAAC 1280

QY 841 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTTCTAACGAAAACATCAAGCGCGTAAT 900

DB 1281 ACTAATGGCTTTACGGCTTCTAGCTAGACATTTTCTAACGAAAACATCAAGCGCGTAAT 1340

QY 901 TTCACTTTTCGAGCAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATCAGCGTTAATT 960

DB 1341 TTCACTTTTCGAGCAACCAAGATATAAGCGCTCGCTGAAATTTGTGAATCAGCGTTAATT 1400

QY 961 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGTTGSCAAAGTGAAGAACGAGGGTGTG 1020

DB 1401 ACTGTCGGTAAAGACGCGCAGTGTAAATCTTATTGTTGSCAAAGTGAAGAACGAGGGTGTG 1460

QY 1021 ATTAGGCTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAAATCACCACGCGAT 1080

DB 1461 ATTAGGCTAAATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAAAATCACCACGCGAT 1520

QY 1081 ATAATAAACCCCAACCATTTACTTTACAGCATTTGCGCGCTGAAATGAAGCGCTCAATCTG 1140

DB 1521 ATAATAAACCCCAACCATTTACTTTACAGCATTTGCGCGCTGAAATGAAGCGCTCAATCTG 1580

QY 1141 GCGCATATTTTTCGCAAGCGGTACATTAATGTCGCTGCTGCCACTATTTTCGAAACCAA 1200

DB 1581 GCGCATATTTTTCGCAAGCGGTACATTAATGTCGCTGCTGCCACTATTTTCGAAACCAA 1640

QY 1201 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTTGTTCTTCGCGC 1260

DB 1641 GGTAAACTTTCTGCTGATTTCTGTAAGCAAGATAAAGCGCAATATTTGTTCTTCGCGC 1700

QY 1261 AAAGAGGTGAAGCGGAAATTTGCGGTGTAAATTTCCGCTCAAAATCAGCAGCTAAAGGC 1320

DB 1701 AAAGAGGTGAAGCGGAAATTTGCGGTGTAAATTTCCGCTCAAAATCAGCAGCTAAAGGC 1760

QY 1321 GCGAGCTGATGATTTACAGCGCATAAAGTCAATTAACAGGTGCGAGTTATCGACCTT 1380

DB 1761 GCGAGCTGATGATTAAGTCCGTAAGTCAATTAACAGGTGCGAGTTATCGACCTT 1820

QY 1381 TCAGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGAGGTAAAGAC 1440

DB 1821 TCAGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGAGGTAAAGAC 1880

QY 1441 GCGATTTCAATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATGATATCAGGC 1500

DB 1881 GCGATTTCAATTTAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATGATATCAGGC 1940

QY 1501 AAAGAAAAGCGGACGCGCTATTGTTGGGCGATATTGCGTTAATTTGACGGCAATATT 1560

DB 1941 AAAGAAAAGCGGACGCGCTATTGTTGGGCGATATTGCGTTAATTTGACGGCAATATT 2000

QY 1561 AAGCGTCAAGGTAGTGTGATATCCGTAAGAACCGGTGGTTTTTGGAGACATCGGGCAT 1620

DB 2001 AAGCGTCAAGGTAGTGTGATATCCGTAAGAACCGGTGGTTTTTGGAGACATCGGGCAT 2060

QY 1621 TATTATCCATTTGACAGCAATGCAATTTGTTGAGGCGATATTGCGTTAATTTGACGGCAATATT 1680

DB 2061 TATTATCCATTTGACAGCAATGCAATTTGTTGAGGCGATATTGCGTTAATTTGACGGCAATATT 2120

QY 1681 GATGTAAAGTGTGACGCGGACGCGCTTTGCGCAATTAATACCGGTATATAATGATGAATTC 1740

DB 2121 GATGTAAAGTGTGACGCGGACGCGCTTTGCGCAATTAATACCGGTATATAATGATGAATTC 2180

QY 1741 CCAACAGCGCCGCTGAAGCAAGCGCTTAAACCAAAATAGGCACTCAAAACAGCGCTA 1800

DB 2181 CCAACAGCGCCGCTGAAGCAAGCGCTTAAACCAAAATAGGCACTCAAAACAGCGCTA 2240

QY 1801 ACCAATACACTATTTCAATTTATCTGAAAACCGCTGGACAAATGAATATAAGCGCATCA 1860

DB 2241 ACCAATACACTATTTCAATTTATCTGAAAACCGCTGGACAAATGAATATAAGCGCATCA 2300

QY 1861 AGAACAATTTACCGTTAATAGTCAATCAACATCGGAAGCAACTCCCACTTAATCTTCAT 1920

Db 2301 AGAAACTTACCCTTAAGCTCAATCAACATCGAAGCAACTCCACATTAATCTCCAT 2360  
QY 1921 AGTAAAGGTCAGCGTGGCGAGGCGTTACAGATTGATGGAGATATTAATCTAAAGGGGA 1980  
Db 2361 AGTAAAGGTCAGCGTGGCGAGGCGTTACAGATTGATGGAGATATTAATCTAAAGGGGA 2420  
QY 1981 AATTTACCAATTTATCTCGCGGATGGGTTGATGTTTCATAAAATATTAAGCTTGATCAG 2040  
Db 2421 AATTTACCAATTTATCTCGCGGATGGGTTGATGTTTCATAAAATATTAAGCTTGATCAG 2480  
QY 2041 GGTTTTAAATATTAACCGCGCTCCGTAGCTTTGAAGGTGGAATAACAAAGCACGC 2100  
Db 2481 GGTTTTAAATATTAACCGCGCTCCGTAGCTTTGAAGGTGGAATAACAAAGCACGC 2540  
QY 2101 GACCGGCAATGCTAAAATTTGCGCCAGGCACTGTAACCAATTAACAGGAGGGAATA 2160  
Db 2541 GACCGGCAATGCTAAAATTTGCGCCAGGCACTGTAACCAATTAACAGGAGGGAATA 2600  
QY 2161 GATTTACGGGCTAACCACTGATCTTTAAACGGAAACGGGTAAAGGTCTGTAATATCATTTCA 2220  
Db 2601 GATTTACGGGCTAACCACTGATCTTTAAACGGAAACGGGTAAAGGTCTGTAATATCATTTCA 2660  
QY 2221 TCAGTGAATTAATTAACCCCAATCTTAGTGCCACAAATTAACATATCTGGGAATATAACA 2280  
Db 2661 TCAGTGAATTAATTAACCCCAATCTTAGTGCCACAAATTAACATATCTGGGAATATAACA 2720  
QY 2281 ATTAACCAACTACGAGAAGAACCTCGTATTGGCAACACAGCCATGATTCGCATGG 2340  
Db 2721 ATTAACCAACTACGAGAAGAACCTCGTATTGGCAACACAGCCATGATTCGCATGG 2780  
QY 2341 AACGTCAGTCTTAACTAGAGACAGGGGCAAAATTTTACCTTTATTAATACATTTCA 2400  
Db 2781 AACGTCAGTCTTAACTAGAGACAGGGGCAAAATTTTACCTTTATTAATACATTTCA 2840  
QY 2401 AGCAATAGCAAGGCTTAACAACACAGTAPAGAAGTCTGAGGGGTGAATTTTAAACGGC 2460  
Db 2841 AGCAATAGCAAGGCTTAACAACACAGTAPAGAAGTCTGAGGGGTGAATTTTAAACGGC 2500  
QY 2461 GTAAATGGCAACATGCTATTCAATCTCAAGAAAGGAGCGAAAGTTAATTTCAAATTAATA 2520  
Db 2901 GTAAATGGCAACATGCTATTCAATCTCAAGAAAGGAGCGAAAGTTAATTTCAAATTAATA 2560  
QY 2521 CCAACGAGAACATGATCAACACAGCAACCTTTTACCAATTCGGTCTGAGCGGATATCA 2580  
Db 2961 CCAACGAGAACATGATCAACACAGCAACCTTTTACCAATTCGGTCTGAGCGGATATCA 3020  
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
Db 3021 GCCACTGGTGGGGCTCTGTTTTTTTGTATATATGCCAACCATTTCTGGCAGAGGGCT 3080  
QY 2641 GAGTTAAAATGATGCAAAATTAATATCTACGGCGCTAATTTTACCTTAATTTCCCAT 2700  
Db 3081 GAGTTAAAATGATGCAAAATTAATATCTACGGCGCTAATTTTACCTTAATTTCCCAT 3140  
QY 2701 GTTCGGGCGATGACGCTTTTAAATCAACAAGACTTAACCAATAATGCAACCAATCA 2760  
Db 3141 GTTCGGGCGATGACGCTTTTAAATCAACAAGACTTAACCAATAATGCAACCAATCA 3200  
QY 2761 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGGTACGACGCAATGCGCATC 2820  
Db 3201 AATTTACGCTCAGACAGCAAGATGATTTTATGACGGGTACGACGCAATGCGCATC 3260  
QY 2821 AATTTACCTACCAATATCCATCTTCGGGCGGTAAATGTCAACCTTGGTGGACAAACTCA 2880  
Db 3261 AATTTACCTACCAATATCCATCTTCGGGCGGTAAATGTCAACCTTGGTGGACAAACTCA 3320  
QY 2881 AGCAGCAGCATTTACGGGGAATATTAATCTAGAGAAAGCAGCAAACTGATAGAGCC 2940  
Db 3321 AGCAGCAGCATTTACGGGGAATATTAATCTAGAGAAAGCAGCAAACTGATAGAGCC 3380  
QY 2941 AATAAGCCCCCTAATCAGCAAAACATTAAGGATAGAGTTAATAAAGCTTGGCAGCTTGCTC 3000  
Db 3381 AATAAGCCCCCTAATCAGCAAAACATTAAGGATAGAGTTAATAAAGCTTGGCAGCTTGCTC 3440

QY 3001 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATGCAGATATTAAGGCAATCTCACATAT 3060  
Db 3441 GTTAATGGGAGTTTAAAGTTTAACTGGCGAAATGCAGATATTAAGGCAATCTCACATAT 3500  
QY 3061 TCAGAAAGCCGACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACACGGCAATTTT 3120  
Db 3501 TCAGAAAGCCGACCTTTTAAAGGAAAGACTAGAGATACCCCTAAATATACACGGCAATTTT 3560  
QY 3121 ACCAATATGGCAGCTCCGCAATTAATATACCAAGAGAGTGGTAAACCTTTGGCAATGTT 3180  
Db 3561 ACCAATATGGCAGCTCCGCAATTAATATACCAAGAGAGTGGTAAACCTTTGGCAATGTT 3620  
QY 3181 ACCAATGATGTTTAAACATTAACACTTACACTTAAACGCAACCAAGAGCATCATC 3240  
Db 3621 ACCAATGATGTTTAAACATTAACACTTACACTTAAACGCAACCAAGAGCATCATC 3680  
QY 3241 GCGGAGATATTAATCAACAAAGAGAGCTTAAATATTACAGACAGTAAATGATGCT 3300  
Db 3681 GCGGAGATATTAATCAACAAAGAGAGCTTAAATATTACAGACAGTAAATGATGCT 3740  
QY 3301 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGAGCTTAAATTTTCCGAT 3360  
Db 3741 GAAATCCAAATTTGGCGCAATATCTCGCAAAAGAGAGCTTAAATTTTCCGAT 3800  
QY 3361 AATAATTAATTCACCAACAGATTAACATCAAAAGGATTTGATGGAGAGCTTAGT 3420  
Db 3801 AATAATTAATTCACCAACAGATTAACATCAAAAGGATTTGATGGAGAGCTTAGT 3860  
QY 3421 TCAGATGCGCAAGTAAATGCCAATTAATTAACCAAAAGAGTAAATTTGACAGAA 3480  
Db 3861 TCAGATGCGCAAGTAAATGCCAATTAATTAACCAAAAGAGTAAATTTGACAGAA 3920  
QY 3481 GACCTAAGTATTTTACAGGTTTCAATTAAGCAGAGATTAACAGCAAAAGTGGTAGAGTTA 3540  
Db 3921 GACCTAAGTATTTTACAGGTTTCAATTAAGCAGAGATTAACAGCAAAAGTGGTAGAGTTA 3980  
QY 3541 ACTATTGGCAACAGTAAATGACGTTAACAGCGGTGCCGAAGCCAAACAGTAACCTTTAAC 3600  
Db 3981 ACTATTGGCAACAGTAAATGACGTTAACAGCGGTGCCGAAGCCAAACAGTAACCTTTAAC 4040  
QY 3601 AATGTTAAAGATTCACAAATCTCTGCTGAGGTCACATGACACTTAATAGCAAGTG 3660  
Db 4041 AATGTTAAAGATTCACAAATCTCTGCTGAGGTCACATGACACTTAATAGCAAGTG 4100  
QY 3661 AARACATCTAGCAGCAATGCGGACGTTAAAGCAATAGCAGCAACGATACCGGCTTAAT 3720  
Db 4101 AARACATCTAGCAGCAATGCGGACGTTAAAGCAATAGCAGCAACGATACCGGCTTAAT 4160  
QY 3721 ATTACTGCAAAATGATAGAGTAAACAAAGATTAATTTCTCTCAAAACAGTAATATC 3780  
Db 4161 ATTACTGCAAAATGATAGAGTAAACAAAGATTAATTTCTCTCAAAACAGTAATATC 4220  
QY 3781 ACCGCTCGGAAAGGTTACACACAGAGGCTCGACCAATTAACGCAACAAATGGCAAA 3840  
Db 4221 ACCGCTCGGAAAGGTTACACACAGAGGCTCGACCAATTAACGCAACAAATGGCAAA 4280  
QY 3841 GCAAGTATTACAAACCAAGGATGATATCAGCGGTACGATTTCCGCTTAACACGTTAGT 3900  
Db 4281 GCAAGTATTACAAACCAAGGATGATATCAGCGGTACGATTTCCGCTTAACACGTTAGT 4340  
QY 3901 GTTAGCGCACTGTTTAAACCACTTAATCCGGCTCAAAATTAAGCGAATCCGGT 3960  
Db 4341 GTTAGCGCACTGTTTAAACCACTTAATCCGGCTCAAAATTAAGCGAATCCGGT 4400  
QY 3961 GAGGCTTAATGATACAGTGCACAGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4020  
Db 4401 GAGGCTTAATGATACAGTGCACAGTACAATTTGGCGGTACAATTTCCGGTAAATACGGTA 4460  
QY 4021 AATGTTACGGCAACGCTGCGGATTTAACAGTGGGAATGGCGAGAAATTAATGGGACA 4080  
Db 4461 AATGTTACGGCAACGCTGCGGATTTAACAGTGGGAATGGCGAGAAATTAATGGGACA 4520

```

Qy 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTCTGACTACTGAAGCCGGTCTTACG 4140
Dy 4521 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTCTGACTACTGAAGCCGGTCTTACG 4580
Qy 4141 ATCACTTCAACTAAGGCTGAGTACCTTCTGGCTCAGAAATGCTAGCATCGCAGGAAGC 4200
Dy 4581 ATCACTTCAACTAAGGCTGAGTACCTTCTGGCTCAGAAATGCTAGCATCGCAGGAAGC 4640
Qy 4201 ATTAATGCTGCTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4260
Dy 4641 ATTAATGCTGCTAATGTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4700
Qy 4261 GATATTAAGCAACCAAGCGGACCTTGGTTATTAAACGCAAAAGATGCTAAAGCTAAATGCT 4320
Dy 4701 GATATTAAGCAACCAAGCGGACCTTGGTTATTAAACGCAAAAGATGCTAAAGCTAAATGCT 4760
Qy 4321 GATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAAGCAAGCGGCTCTGGTACTGTG 4380
Dy 4761 GATGATCAGGTGATAGTACAGAGTGAATGCAAGTCAAGCAAGCGGCTCTGGTACTGTG 4820
Qy 4381 ACTGGCGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGGTTA 4440
Dy 4821 ACTGGCGCAACCTCAAGCAGTGTGAATATCACTGGGGATTTAAACACAGTAATGGGTTA 4880
Qy 4441 ATATATCTTTCGAAGATGTAGAAACACTGTGCGCTTAAAGAGCAAGAAATTTAGAGTG 4500
Dy 4881 ATATATCTTTCGAAGATGTAGAAACACTGTGCGCTTAAAGAGCAAGAAATTTAGAGTG 4940
Qy 4501 AAATATATCCAGCAGGTGTACAGAGTGTAGAGAGTAAATTTGAAGCGAAACCGGCTCTT 4560
Dy 4941 AAATATATCCAGCAGGTGTACAGAGTGTAGAGAGTAAATTTGAAGCGAAACCGGCTCTT 5000
Qy 4561 GAAAGTAAAGATTTATCTGATGAAGAGAGAAACACTTACGCTTAACTTGGTGTAACT 4620
Dy 5001 GAAAGTAAAGATTTATCTGATGAAGAGAGAAACACTTACGCTTAACTTGGTGTAACT 5060
Qy 4621 GCTGTACGTTTGTGAGCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4680
Dy 5061 GCTGTACGTTTGTGAGCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5120
Qy 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGGTGTCTCAAGTGGTAAAT 4740
Dy 5121 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGGTAAAGCGGTGTCTCAAGTGGTAAAT 5180
Qy 4741 GCGCACAGATGTACCAATGCTGTCAGCAGTGGACAGCGCTAGTCAAGTAATTTGACAG 4800
Dy 5181 GCGCACAGATGTACCAATGCTGTCAGCAGTGGACAGCGCTAGTCAAGTAATTTGACAG 5240
Qy 4801 GTAGATTTTCATCTGCAATGAAGTCATTTATTTTTCGTAATTTTACTGTGTGGGTTAA 4860
Dy 5241 GTAGATTTTCATCTGCAATGAAGTCATTTATTTTTCGTAATTTTACTGTGTGGGTTAA 5300
Qy 4861 GTTCAGTACGGGCTTTACCCATCTTGTAAAGATTTACGAGAGTAATCAATGAATTTT 4920
Dy 5301 GTTCAGTACGGGCTTTACCCATCTTGTAAAGATTTACGAGAGTAATCAATGAATTTT 5360
Qy 4921 AACAGGTTATTATTATG 4937
Dy 5361 AACAGGTTATTATTATG 5377

```

## RESULT 12

```

US-08-469-880-6
; Sequence 6, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

```

; STREET: Bldg. 1
; City: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beckstesser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-6

```

```

Query Match 97.6%; Score 4820.2; DB 2; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

```

Qy 1 TAAATATACAGAGATAATAAAATAAATCAAGATTTTGTGATGACAAACAAATTTACAA 60
Dy 441 TAAATATACAGAGATAATAAAATAAATCAAGATTTTGTGATGACAAACAAATTTACAA 500
Qy 61 CACCTTTTGGCAGTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCCATATAA 120
Dy 501 CACCTTTTGGCAGTCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCCATATAA 560
Qy 121 AATGATATAATCTTTTCATCTTTTCAATCTTTTAACTTTTCATCTTTTCATCTTTTCAT 180
Dy 561 AATGATATAATCTTTTCATCTTTTCAATCTTTTAACTTTTCATCTTTTCATCTTTTCAT 620
Qy 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240
Dy 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680
Qy 241 CACATGAATATGTAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Dy 681 CACATGAATATGTAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
Qy 301 GAACGCAATGATAAAGTAAATTTTAACTTCACTTAACTTCACTTAACTTCACTTAACTT 360
Dy 741 GAACGCAATGATAAAGTAAATTTTAACTTCACTTAACTTCACTTAACTTCACTTAACTT 800
Qy 361 ATATATCTCTCAAAATTCAGCAAAACCGCTTAACTTAACTTAACTTAACTTAACTTAACT 420
Dy 801 ATATATCTCTCAAAATTCAGCAAAACCGCTTAACTTAACTTAACTTAACTTAACTTAACT 860
Qy 421 CGGGGTGTCACCATTTCCACAGAAAAGGCTTCCGCTATGTTACTTACTTCTTACTTACT 480

```

Db 861 CGGGGTTGTGACCAATTCACAGAAAAGGCGAGGAAAACCTGCTCGCATGAAAGTGGCT 920  
QY 481 CACTTAGCGTTAAAGCCACTTTCGGCTATGTTACTATCTTTTAGGTGTACATCTATTTC 540  
Db 921 CACTTAGCGTTAAAGCCACTTTCGGCTATGTTACTATCTTTTAGGTGTACATCTATTTC 980  
QY 541 CAATCTGTTTGTAGCAAGCGCTTACAGGAATGGATGTAGTACAGCGGCACAGCCACTAT 600  
Db 981 CAATCTGTTTGTAGCAAGCGCAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTAT 1040  
QY 601 CAAGTAGATGTTAATAAAGCAATTTCCGCAACAGTGTGTGAGCTATCATTAATTTGAAA 660  
Db 1041 AAGAAAACAAGTAAATAAAGCAATTTCCGCAACAGTGTGTGAGCTATCATTAATTTGAAA 1100  
QY 661 CAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTCAGAAAACAACCAACTCCGCC 720  
Db 1101 CAATTTAAACATCGACCAAAATGAAATGGTGCAGTTTTCAGAAAACAACCAACTCCGCC 1160  
QY 721 GTATTCAGCGCTGTATCATCTACCAAAATCTCCCAATTTAAAGGGATTTTAGATTTCAAC 780  
Db 1161 GTATTCAGCGCTGTATCATCTACCAAAATCTCCCAATTTAAAGGGATTTTAGATTTCAAC 1220  
QY 781 GGACAGCTCTTTTAATCAACCAATGGTATCACAAATAGGTAAAGAGCAATTTATTAAC 840  
Db 1221 GGACAGCTCTTTTAATCAACCAATGGTATCACAAATAGGTAAAGAGCAATTTATTAAC 1280  
QY 841 ACTAATGGCTTTAGCGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGGCGCGTAAT 900  
Db 1281 ACTAATGGCTTTAGCGCTTCTAGCTAGACATTTCTAACGAAAACATCAAGGCGCGTAAT 1340  
QY 901 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCACGGTTTAAT 960  
Db 1341 TTCACCTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTAATCACGGTTTAAT 1400  
QY 961 ACTGTCGGTAAAGACGCGAGTAAATCTTATTTGGTGGCAAGTGAAGAACGAGGCTGTG 1020  
Db 1401 ACTGTCGGTAAAGACGCGAGTAAATCTTATTTGGTGGCAAGTGAAGAACGAGGCTGTG 1460  
QY 1021 ATTAGCTTAATGTGGCAGCATTTCTTACTCGAGGGCAAAAATCACCATCAGCGAT 1080  
Db 1461 ATTAGCTTAATGTGGCAGCATTTCTTACTCGAGGGCAAAAATCACCATCAGCGAT 1520  
QY 1081 ATATAACCCCAACATTTACTACAGCATTTGCCGCTGAAATGAAGCGGTCAATCTG 1140  
Db 1521 ATATAACCCCAACATTTACTACAGCATTTGCCGCTGAAATGAAGCGGTCAATCTG 1580  
QY 1141 GCGGATATTTTGGCAAGGCGGTAAATTAATGTCGCTGCTGCCACTATTCGAAACCAA 1200  
Db 1581 GCGGATATTTTGGCAAGGCGGTAAATTAATGTCGCTGCTGCCACTATTCGAAACCAA 1640  
QY 1201 GGTAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAAGCGGCAATATTTGTTCTTCGCC 1260  
Db 1641 GGTAACTTTCTGCTGATTTCTGTAAAGCAAGATAAAAGCGGCAATATTTGTTCTTCGCC 1700  
QY 1261 AAAGAGGTTGAAGCGGAAATTTGGCGGTGAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1320  
Db 1701 AAAGAGGTTGAAGCGGAAATTTGGCGGTGAATTTCCGCTCAAAATCAGCAAGCTTAAAGGC 1760  
QY 1321 GCGAGCTGATGATTAACAGCGGATAAAGTACATTTAAACAGCTGCGAGTTATCGACTTT 1380  
Db 1761 GCGAGCTGATGATTAACAGCTGATAAAGTACATTTAAACAGCTGCGAGTTATCGACTTT 1820  
QY 1381 TCAGTTAAGAGGGGGAACCTTTACCTTGGCGGTGACGAGCGGCGGAGGTAAAGAAC 1440  
Db 1821 TCAGTTAAGAGGGGGAACCTTTACCTTGGCGGTGACGAGCGGCGGAGGTAAAGAAC 1880  
QY 1441 GCGATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1500  
Db 1881 GCGATTCATTAAGCAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGC 1940  
QY 1501 AAAGAAAAGCGGAGCGCTATTGTGGGCGGATATTTGGCTTAATTTGACGCGCAATAT 1560  
Db 1941 AAAGAAAAGCGGAGCGCTATTGTGGGCGGATATTTGGCTTAATTTGACGCGCAATAT 2000

QY 1561 AACGCTCAAGCTAGTGTGATATCGCTAAAAACCGTGGTGTGTTGGAGACATCGGGCAT 1620  
Db 2001 AACGCTCAAGCTAGTGTGATATCGCTAAAAACCGTGGTGTGTTGGAGACATCGGGCAT 2060  
QY 1621 TATTTATCCATTTGACAGCAATGCTTTTAAACAAGAGTGTGCTAGACCCCTGAT 1680  
Db 2061 TATTTATCCATTTGACAGCAATGCTTTTAAACAAGAGTGTGCTAGACCCCTGAT 2120  
QY 1681 GATGTAACAATTTGAAGCGGAAGACCCCTTCGCAATATATACGGGTATTAATGATGATTC 1740  
Db 2121 GATGTAACAATTTGAAGCGGAAGACCCCTTCGCAATATATACGGGTATTAATGATGATTC 2180  
QY 1741 CCAACAGGCGCGGTGAAGCAAGCGACCTTAAAAAATAGGGAACCTCAAAACAACGGTA 1800  
Db 2181 CCAACAGGCGCGGTGAAGCAAGCGACCTTAAAAAATAGGGAACCTCAAAACAACGGTA 2240  
QY 1801 ACCAATACAACATTTCTCAAAATTTATCTGAAAAACGCTTGCAATGAATATAACGCAATCA 1860  
Db 2241 ACCAATACAACATTTCTCAAAATTTATCTGAAAAACGCTTGCAATGAATATAACGCAATCA 2300  
QY 1861 AGAAAACCTTACCGTTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCTCCAT 1920  
Db 2301 AGAAAACCTTACCGTTTAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCTCCAT 2360  
QY 1921 AGTAAAGTTCAGCTGGCGGAGCGTTCAGATTGATGGAGATATTACTTTAAAGCGCGA 1980  
Db 2361 AGTAAAGTTCAGCTGGCGGAGCGTTCAGATTGATGGAGATATTACTTTAAAGCGCGA 2420  
QY 1981 AATTTAAACATTTATCTGGCGGATGGTGTGATTTTCATATAAATATTACGGTTGATCAG 2040  
Db 2421 AATTTAAACATTTATCTGGCGGATGGTGTGATTTTCATATAAATATTACGGTTGATCAG 2480  
QY 2041 GGTTTTTTAAATATTACCGCCCTTCCGCTAGCTTTTGAAGGTGGAATTAACAAAGCAACGC 2100  
Db 2481 GGTTTTTTAAATATTACCGCCCTTCCGCTAGCTTTTGAAGGTGGAATTAACAAAGCAACGC 2540  
QY 2101 GACGCGGCAAAATGCTAAAAATTTGTCGCCAGGCACTGTAAACCATTTACAGGAGAGGGA 2160  
Db 2541 GACGCGGCAAAATGCTAAAAATTTGTCGCCAGGCACTGTAAACCATTTACAGGAGAGGGA 2600  
QY 2161 GATTTTCAGGCTTAAACAACGATATCTTTAAACGGAACGGGTAAAAGGTCTGAATATCATTTCA 2220  
Db 2601 GATTTTCAGGCTTAAACAACGATATCTTTAAACGGAACGGGTAAAAGGTCTGAATATCATTTCA 2660  
QY 2221 TCAGTCAATAATTTAAACCCACATCTTAGTGGCACAATTTAAACATATCTGGGAATATAACA 2280  
Db 2661 TCAGTCAATAATTTAAACCCACATCTTAGTGGCACAATTTAAACATATCTGGGAATATAACA 2720  
QY 2281 ATTTAAACCAACTACGAGAAGACACCTCGTATTGGCAACCGCATGATTCGCACTGG 2340  
Db 2721 ATTTAAACCAACTACGAGAAGACACCTCGTATTGGCAACCGCATGATTCGCACTGG 2780  
QY 2341 AACGTCAGTGTCTTAACTTAGAGAGCGGCAATTTTACCTTTTAAATACATTTCA 2400  
Db 2781 AACGTCAGTGTCTTAACTTAGAGAGCGGCAATTTTACCTTTTAAATACATTTCA 2840  
QY 2401 AGCAATAGCAAGGCTTAAACAACAGTATAGAAAGTCTGCGAGGGTGAATTTTAAACGGC 2460  
Db 2841 AGCAATAGCAAGGCTTAAACAACAGTATAGAAAGTCTGCGAGGGTGAATTTTAAACGGC 2900  
QY 2461 GTAAATGGCAACATGTCTCAATCTCAATCTCAAAAGAGGAGCAAGTAAATTTCAATTTAAA 2520  
Db 2901 GTAAATGGCAACATGTCTCAATCTCAATCTCAAAAGAGGAGCAAGTAAATTTCAATTTAAA 2960  
QY 2521 CCAACGAGAACATGTAACAACAAGCAACCTTTTACCAATTCGGTTTTTTAGCCAATATACACA 2580  
Db 2961 CCAACGAGAACATGTAACAACAAGCAACCTTTTACCAATTCGGTTTTTTAGCCAATATACACA 3020  
QY 2581 GCCACTGGTGGGCTCTGTTTTTTTATATATATCCCAACCATTTCTGCGAGAGGGCT 2640  
Db 3021 GCCACTGGTGGGCTCTGTTTTTTTATATATATCCCAACCATTTCTGCGAGAGGGCT 3080





Db 5241 GTAGATTTTCATCCTGCAATGAAGTCATTTATTTTCGTATTTTACTGTGGTTAAA 5300  
 QY 4861 GTTCAGTAGGGCTTTACCATCTTGTAAAAAATACGAGAAATACAAATAAAGTATTTT 4920  
 Db 5301 GTTCAGTAGGGCTTTACCATCTTGTAAAAAATACGAGAAATACAAATAAAGTATTTT 5360  
 QY 4921 AACAGGTTATTTATG 4937  
 Db 5361 AACAGGTTATTTATG 5377

## RESULT 13

US-08-728-470-6  
 ; Sequence 6, Application US/08728470  
 ; Patent No. 5928651  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barenkamp, Stephen J  
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/728.470  
 ; FILING DATE:  
 ; CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/302,832  
 ; FILING DATE: 16-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9205704.1

ATTORNEY/AGENT INFORMATION:  
 ; NAME: Berkstesser, Jerry W  
 ; REGISTRATION NUMBER: 22,654  
 ; REFERENCE/DOCKET NUMBER: 1038-633  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 415-0810  
 ; TELEFAX: (703) 415-0813  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9323 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 US-08-728-470-6

Query Match  
 Best Local Similarity 97.6%; Score 4820.2; DB 2; Length 9323;  
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAAAGTATAAATAAATCAAGATTTTGTGATGACAAACAAATACAA 60  
 Db 441 TAAATATACAAAGTATAAATAAATCAAGATTTTGTGATGACAAACAAATACAA 500

QY 61 CACCTTTTTCAGTCTATATGCAATATTTTAAAAAATAGTATATAATCCGCCATATA 120  
 Db 501 CACCTTTTTCAGTCTATATGCAATATTTTAAAAAATAGTATATAATCCGCCATATA 560

QY 121 AATGGTATATATCTTTCATCTTTTCATCTTTTAATCTTTTCATCTTTTCATCTTTTCAT 180  
 Db 561 AATGGTATATATCTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCAT 620  
 QY 181 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 240  
 Db 621 CTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 680  
 QY 241 CACATGAATATGTAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300  
 Db 681 CACATGAATATGTAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 740  
 QY 301 GAACGCAATGATATAAGTAAATTTAATTTGTTCAACTAACCTTAGGAGGAGGAGGAGG 360  
 Db 741 GAACGCAATGATATAAGTAAATTTAATTTGTTCAACTAACCTTAGGAGGAGGAGGAGG 800  
 QY 361 ATATATCGTCTCAAAATTCAGCAAAACGCTTGAATGCTTTGGTTCGCTGCTGCTGCTG 420  
 Db 801 ATATATCGTCTCAAAATTCAGCAAAACGCTTGAATGCTTTGGTTCGCTGCTGCTGCTG 860  
 QY 421 CGGGTTGTGACCATTCACAGAAAAGGCTTCCGCTATGTTACTATCTTTAGSTGTAAC 480  
 Db 861 CGGGTTGTGACCATTCACAGAAAAGGCTTCCGCTATGTTACTATCTTTAGSTGTAAC 920  
 QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGSTGTAACATCTATTCCA 540  
 Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATGTTACTATCTTTAGSTGTAACATCTATTCCA 980  
 QY 541 CAATCTGTTTTAGCAAGCGCTTACAAAGAAATGGATGTAGTACACGGCACAGCCTATG 600  
 Db 981 CAATCTGTTTTAGCAAGCGCTTAAATTAACATCGACCAAAATGAAATGGTGCAGTTTTTAC 1040  
 QY 601 CAAGTAGATGGTAAATAAACCCTATATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 660  
 Db 1041 AAGAAAACAAGTAAATAAACCCTATATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 1100  
 QY 661 CAATTTAATCATCGACCAAAATGAAATGGTGCAGTTTTTACAGAAAACAACAACTCCGCC 720  
 Db 1101 CAATTTAATCATCGACCAAAATGAAATGGTGCAGTTTTTACAGAAAACAACAACTCCGCC 1160  
 QY 721 GTATTTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGGATTTTATAGTTTAA 780  
 Db 1161 GTATTTCAACCGTGTATACATCTAACCAATCTCCCAATTTAAAGGGATTTTATAGTTTAA 1220  
 QY 781 GGACAGTCTTTTAAATCAACCAATGGTATCAGTATCAGTAAAGCAATATTAAC 840  
 Db 1221 GGACAGTCTTTTAAATCAACCAATGGTATCAGTATCAGTAAAGCAATATTAAC 1280  
 QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGGCGCGTAAT 900  
 Db 1281 ACTAATGGCTTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGGCGCGTAAT 1340  
 QY 901 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTAATCAAGGTTTAAAT 960  
 Db 1341 TTCACCTTCGAGCAAAACCAAGATATAAGCGCTCGCTGAAATTTGTAATCAAGGTTTAAAT 1400  
 QY 961 ACTGTCCGTTAAGACGCGCTGTAAATCTTTATTTGGTGGCAAGTGAAGGAGGTTG 1020  
 Db 1401 ACTGTCCGTTAAGACGCGCTGTAAATCTTTATTTGGTGGCAAGTGAAGGAGGTTG 1460  
 QY 1021 ATTTAGCGTAAATGGTGCAGCATTTCTTTACTCGCGGGCAAAAAATCACCATCAGCGAT 1080  
 Db 1461 ATTTAGCGTAAATGGTGCAGCATTTCTTTACTCGCGGGCAAAAAATCACCATCAGCGAT 1520  
 QY 1081 ATAAATACCCCAACCATTTACTTACGATTTGCCGCGCTGAAAATGAAGCGGTCAATCTG 1140  
 Db 1521 ATAAATACCCCAACCATTTACTTACGATTTGCCGCGCTGAAAATGAAGCGGTCAATCTG 1580  
 QY 1141 GCGGATATTTTTCGCAAAAGCGGTAACATTAATTTGCGCTGCTGCCACTATTTCCGAAACCAA 1200  
 Db 1581 GCGGATATTTTTCGCAAAAGCGGTAACATTAATTTGCGCTGCTGCCACTATTTCCGAAACCAA 1640



us-10-092-880-3.rni

Tue Mar 25 08:39:01 2003

Db 3801 AAAATTAATATACCAACACAGATAAACAATCAAAAAGGGTATTGATGAGAGGAGCTCTAGT 3860  
 Qy 3421 TCAGATGCGCAAGTAAATGCAACCTAACTATTAAACACCAAGAAATTTGAAATTTGACAGAA 3480  
 Db 3861 TCAGATGCGCAAGTAAATGCAACCTAACTATTAAACACCAAGAAATTTGAAATTTGACAGAA 3920  
 Qy 3481 GACCTAAGTATTTAGGTTTCAATTAAGACAGAGATTACAGCAAGAGATGCTAGAGATTTA 3540  
 Db 3921 GACCTAAGTATTTAGGTTTCAATTAAGACAGAGATTACAGCAAGAGATGCTAGAGATTTA 3980  
 Qy 3541 ACTATTGGCAACAGTAATGAGGTAAACAGCGGTGCGGAGCGCAACAGCAAGTAACCTTTTAAAC 3600  
 Db 3981 ACTATTGGCAACAGTAATGAGGTAAACAGCGGTGCGGAGCGCAACAGCAAGTAACCTTTTAAAC 4040  
 Qy 3601 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAATATGACACTTAATAGCAAGTG 3660  
 Db 4041 AATGTTAAAGATTCAAAAATCTCTGCTGACGGTCAATATGACACTTAATAGCAAGTG 4100  
 Qy 3661 AAAACATCTAGCAGCAATGCGGAGCGTGAAGCAATAGCGCAACAGGATACCGGCTTAACT 3720  
 Db 4101 AAAACATCTAGCAGCAATGCGGAGCGTGAAGCAATAGCGCAACAGGATACCGGCTTAACT 3780  
 Qy 3721 ATTACTGCAAAAATCTAGAAATTAACAAAGATATTACTTCTCTCAAAACAGTAAATATC 3780  
 Db 4161 ATTACTGCAAAAATCTAGAAATTAACAAAGATATTACTTCTCTCAAAACAGTAAATATC 4220  
 Qy 3781 ACCGCTGCGAAAGGTTACCAACAGCAGGCTGACCAATTAAGCGCAACAAATGGCAAA 3840  
 Db 4221 ACCGCTGCGAAAGGTTACCAACAGCAGGCTGACCAATTAAGCGCAACAAATGGCAAA 4280  
 Qy 3841 GCAAGTATTACACCAACAGGATGATATCAGCGGTACGATTTCCGCTAACACGTAAGT 3900  
 Db 4281 GCAAGTATTACACCAACAGGATGATATCAGCGGTACGATTTCCGCTAACACGTAAGT 4340  
 Qy 3901 GTTAGCGGACTGGTATTTAAACCACTAAATCCGCTCAAAAATTTGAAGCGAAATCGGGT 3960  
 Db 4341 GTTAGCGGACTGGTATTTAAACCACTAAATCCGCTCAAAAATTTGAAGCGAAATCGGGT 4400  
 Qy 3961 GAGGCTAATGTAACAGTGCACAGGTACAAATGGCGGTACAAATTTCCGCTAATACGTA 4020  
 Db 4401 GAGGCTAATGTAACAGTGCACAGGTACAAATGGCGGTACAAATTTCCGCTAATACGTA 4460  
 Qy 4021 AATGTTACGGCAACCGCTGGCGATTTAAACAGTTGGGATGCGGCAAGATTAATGCGACA 4080  
 Db 4461 AATGTTACGGCAACCGCTGGCGATTTAAACAGTTGGGATGCGGCAAGATTAATGCGACA 4520  
 Qy 4081 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGCTTCTAGC 4140  
 Db 4521 GAAGGAGCTGCAACCTTAACCGCAACAGGGAATACCTTGACTACTGAAGCCGCTTCTAGC 4580  
 Qy 4141 ATCACTTCAACTAAGGGTCAAGTACCTCTTGGCTCAGAAATGATGATGATGATGATGATG 4200  
 Db 4581 ATCACTTCAACTAAGGGTCAAGTACCTCTTGGCTCAGAAATGATGATGATGATGATGATG 4540  
 Qy 4201 ATTAATGCTGCTAATGTCATTAATTAATACAGGCACTTTAACCCCGTGGCAGGCTCG 4260  
 Db 4641 ATTAATGCTGCTAATGTCATTAATTAATACAGGCACTTTAACCCCGTGGCAGGCTCG 4700  
 Qy 4261 GATATTAAAGCAACAGCGGCACTTGGTTATTAAACGCAAAAGATGCTAAGCTAAATGGT 4320  
 Db 4701 GATATTAAAGCAACAGCGGCACTTGGTTATTAAACGCAAAAGATGCTAAGCTAAATGGT 4760  
 Qy 4321 GATGCTACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4380  
 Db 4761 GATGCTACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4820  
 Qy 4381 ACTGCGCAACCTCAAGCAGTGTGATATCACTGGGATTTAAACACAGTAAATGGTTA 4440  
 Db 4821 ACTGCGCAACCTCAAGCAGTGTGATATCACTGGGATTTAAACACAGTAAATGGTTA 4880  
 Qy 4441 AATATCATTTTCAAGAGTGGTAGAACAACACTGTGCGCTTAAAGAGGCAAGAAATTCAGCTG 4500  
 Db 4881 AATATCATTTTCAAGAGTGGTAGAACAACACTGTGCGCTTAAAGAGGCAAGAAATTCAGCTG 4940

Qy 4501 AAATATATCCAGCCAGCTGTAGCAAGTGTAGAAAGCTAAATTTGAAGCGAAACGCTCCTT 4560  
 Db 4941 AAATATATCCAGCCAGCTGTAGCAAGTGTAGAAAGCTAAATTTGAAGCGAAACGCTCCTT 5000  
 Qy 4561 GAAAAGTAAAGATTTATCTGATGAAAGAGAGAAACATTAGCTAAACTTGGTGTAAAGT 4620  
 Db 5001 GAAAAGTAAAGATTTATCTGATGAAAGAGAGAAACATTAGCTAAACTTGGTGTAAAGT 5060  
 Qy 4621 GCTGTACGTTTTGTGAGCCAAATTAATACAAATTAATACAAATTAATACAAATTAATAC 4680  
 Db 5061 GCTGTACGTTTTGTGAGCCAAATTAATACAAATTAATACAAATTAATACAAATTAATAC 5120  
 Qy 4681 ACCAGACCGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 4740  
 Db 5121 ACCAGACCGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 5180  
 Qy 4741 GCGCCACGAGTATGATACCAATGTTGCTGACGATGACAGCGGTAGTCAAGTCAAGTCAAG 4800  
 Db 5181 GCGCCACGAGTATGATACCAATGTTGCTGACGATGACAGCGGTAGTCAAGTCAAGTCAAG 5240  
 Qy 4801 GTAGATTTTCATCTCTGCAATGAAATGTTATTTTTCGTTATTTTTCGTTATTTTTCGTT 4860  
 Db 5241 GTAGATTTTCATCTCTGCAATGAAATGTTATTTTTCGTTATTTTTCGTTATTTTTCGTT 5300  
 Qy 4861 GTTCAGTACGGCTTTTACCCATCTTCTGTAATAAATTTACGAGAAATACGAGAAATTTT 4920  
 Db 5301 GTTCAGTACGGCTTTTACCCATCTTCTGTAATAAATTTACGAGAAATACGAGAAATTTT 5360  
 Qy 4921 AACAGGTTATTATTATG 4937  
 Db 5361 AACAGGTTATTATTATG 5377

RESULT 14  
 US-08-617-697-6  
 ; Sequence 6, Application US/08617697  
 ; Patent No. 5977336  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barenkamp, Stephen J  
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
 ; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.  
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202-0286  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/617,697  
 ; FILING DATE: 01-APR-1996  
 ; CLASSIFICATION: 424  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/302,832  
 ; FILING DATE: 05-OCT-1994  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US PCT/US93/02166  
 ; FILING DATE: 16-MAR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Berkstresser, Jerry W  
 ; REGISTRATION NUMBER: 22,651  
 ; REFERENCE/DOCKET NUMBER: 1038-557  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 415-0810  
 ; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9323 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-617-697-6

Query Match 97.6%; Score 4820.2; DB 2; Length 9323;  
 Best Local Similarity 98.5%; Pred. No. 0;  
 Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy	1	TAATATACAGATATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA	60
Db	441	TAATATACAGATATAAATAAATCAAGATTTTGTGATGACAAACAATAATACAA	500
Qy	61	CACCTTTTTCAGCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCCATATA	120
Db	501	CACCTTTTTCAGCTATATGCAAAATATTTAAAAAATAGTATAAATCCGCCATATA	560
Qy	121	ATGGTATAATCTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCT	180
Db	561	ATGGTATAATCTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCT	620
Qy	181	CTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCT	240
Db	621	CTTTCATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCTTTTCAATCT	680
Qy	241	CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	300
Db	681	CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	740
Qy	301	GAACGCAATGATAAGTAATTTAATTTGTTCACTTAACCTTAGGAGAAATATGAACA	360
Db	741	GAACGCAATGATAAGTAATTTAATTTGTTCACTTAACCTTAGGAGAAATATGAACA	800
Qy	361	ATATATCGTCTCAATTCAGCAACCCCTGAATGCTTTGGTGTCTGTCTGAATTTG	420
Db	801	ATATATCGTCTCAATTCAGCAACCCCTGAATGCTTTGGTGTCTGTCTGAATTTG	480
Qy	421	CGGGTGTGACCAATCCAGAGAAAGGCTTCGGCTATGTATCTATCTTTAGTGTAC	480
Db	861	CGGGTGTGACCAATCCAGAGAAAGGCTTCGGCTATGTATCTATCTTTAGTGTAC	920
Qy	481	CACCTAGGCTTAAAGCCACTTTCCGCTATGTATCTATCTTTAGTGTACCAATTC	540
Db	921	CACCTAGGCTTAAAGCCACTTTCCGCTATGTATCTATCTTTAGTGTACCAATTC	980
Qy	541	CAATCTGTTTACCAAGCGCTTACAGGAATGGATGTAGTACAGGCACGACCTATG	600
Db	981	CAATCTGTTTACCAAGCGCTTACAGGAATGGATGTAGTACAGGCACGACCTATG	1040
Qy	601	CAAGTATAGGTATATAAACCAATATCCGCAACAGTGTGACGTATCATTAATTTG	660
Db	1041	CAAGTATAGGTATATAAACCAATATCCGCAACAGTGTGACGTATCATTAATTTG	1100
Qy	661	CAATTTAAACATCCAGCAAAATGAATGTCAGTGTTTTACAGAAACAACCACTCC	720
Db	1101	CAATTTAAACATCCAGCAAAATGAATGTCAGTGTTTTACAGAAACAACCACTCC	1160
Qy	721	GTATTCACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTAGATTTAC	780
Db	1161	GTATTCACCGTGTACATCTAACCAATCTCCCAATTTAAAGGGATTTTAGATTTAC	1220
Qy	781	GGCAAGTCTTTTAAATCAACCCAAATGGTATCAATAGTAAAGACCAATTAATTA	840
Db	1221	GGCAAGTCTTTTAAATCAACCCAAATGGTATCAATAGTAAAGACCAATTAATTA	1280
Qy	841	ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAACATCAAGCCGCT	900
Db	1281	ACTAATGGCTTTACGGCTTCTAGCTAGACATTTCTAACGAAACATCAAGCCGCT	1340

Qy	901	TTACCTTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTGAATCAGGTTAAT	960
Db	1341	TTACCTTTCGAGCAAAACCAAGATTAAGCGCTCGCTGAAATTTGTGAATCAGGTTAAT	1400
Qy	961	ACTGTGCGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGAGGAGGTG	1020
Db	1401	ACTGTGCGTAAAGACGCGAGTGTAAATCTTATTTGGTGGCAAGTGAAGAGGAGGTG	1460
Qy	1021	ATTACGCTAAATGGTGGCAGCATTTCTTTTACTCGCAGGCAAAATAACCATCAGCAT	1080
Db	1461	ATTACGCTAAATGGTGGCAGCATTTCTTTTACTCGCAGGCAAAATAACCATCAGCAT	1520
Qy	1081	ATAATAACCAACCACTTACTTACAGCATTTCCCGCCTGAAATGAAGCGGTCAATCTG	1140
Db	1521	ATAATAACCAACCACTTACTTACAGCATTTCCCGCCTGAAATGAAGCGGTCAATCTG	1580
Qy	1141	GGCGATATTTTCCCAAAAGCGGTAAACATTAATGTCCGTGCGCACTATTCGAAACCA	1200
Db	1581	GGCGATATTTTCCCAAAAGCGGTAAACATTAATGTCCGTGCGCACTATTCGAAACCA	1640
Qy	1201	GGTAAACTTCTCTGATTTCTGTAAAGCAAGATAAAGCGCAATATTTCTTTCCGCC	1260
Db	1641	GGTAAACTTCTCTGATTTCTGTAAAGCAAGATAAAGCGCAATATTTCTTTCCGCC	1700
Qy	1261	AAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAG	1320
Db	1701	AAAGAGGTGAAGCGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTAAAG	1760
Qy	1321	GGCAAGCTCATGATTTACAGCGCATTAAGTCACATTTAAACAGGTGCACTTATCGAC	1380
Db	1761	GGCAAGCTCATGATTTACAGCGCATTAAGTCACATTTAAACAGGTGCACTTATCGAC	1820
Qy	1381	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGAGGTAAAC	1440
Db	1821	TCAGGTAAAGAGGGGAGAACTTACCTTGGCGGTGACGAGCGCGGAGAGGTAAAC	1880
Qy	1441	GGCATTCATTTAGCAAAAGAAACCTTTTACGAGGAGGAGGAGGAGGAGGAGGAG	1500
Db	1881	GGCATTCATTTAGCAAAAGAAACCTTTTACGAGGAGGAGGAGGAGGAGGAGGAG	1940
Qy	1501	AAAGAAAGGCGGAGCGCTATTTGTGGGGGATTTTGGGTTTAAATGACGCAATAT	1560
Db	1941	AAAGAAAGGCGGAGCGCTATTTGTGGGGGATTTTGGGTTTAAATGACGCAATAT	2000
Qy	1561	AAAGAAAGGCGGAGCGCTATTTGTGGGGGATTTTGGGTTTAAATGACGCAATAT	1620
Db	2001	AAAGAAAGGCGGAGCGCTATTTGTGGGGGATTTTGGGTTTAAATGACGCAATAT	2060
Qy	1621	TATTTATCCATTGACAGCAATGCAATTTTAAACAAAGAGAGTGGTGTAGACCTGAT	1680
Db	2061	TATTTATCCATTGACAGCAATGCAATTTTAAACAAAGAGAGTGGTGTAGACCTGAT	2120
Qy	1681	CATGTAAATTTGAAGCGAAGACCCCTTTCGCAATAATACCGGTATAATGATGAATTC	1740
Db	2121	CATGTAAATTTGAAGCGAAGACCCCTTTCGCAATAATACCGGTATAATGATGAATTC	2180
Qy	1741	CCACAGGCAACCGGTGAAGCAAGCGCCCTTAAACAAAGAGAGTGGTGTAGACCTGAT	1800
Db	2181	CCACAGGCAACCGGTGAAGCAAGCGCCCTTAAACAAAGAGAGTGGTGTAGACCTGAT	2240
Qy	1801	ACCAATACCACTTATTTCAATTTATCTGAAACACCGCTTAAACAAAGAGAGTGGT	1860
Db	2241	ACCAATACCACTTATTTCAATTTATCTGAAACACCGCTTAAACAAAGAGAGTGGT	2300
Qy	1861	AGAAACTTACCGTATTAATAGCTCAATCAATCAGCAAGCGCTGGAATTAATACGCA	1920
Db	2301	AGAAACTTACCGTATTAATAGCTCAATCAATCAGCAAGCGCTGGAATTAATACGCA	2360
Qy	1921	AGTAAAGGTTCAGCGTGGCGAGCGCTTACAGATTCATGAGATATTACTTCTAAAG	1980
Db	2361	AGTAAAGGTTCAGCGTGGCGAGCGCTTACAGATTCATGAGATATTACTTCTAAAG	2420
Qy	1981	AATTTAACCACTTATTTCTGCGGATGGGTTGATGCTTCAATAAAATATTTACGCT	2040



Db 2421 AATTTAAACCAATTTATCTGGCGGATGGTTGATGTTTCATATAAATATTACGGCTTGATCAG 2480  
QY 2041 GGTFTTTTAAATATTACCGCGCTTCGGTAGCTTTTCAAGTGGAAATACAAAGCAGCG 2100  
Db 2481 GGTFTTTTAAATATTACCGCGCTTCGGTAGCTTTTGAAGTGGAAATACAAAGCAGCG 2540  
QY 2101 GACCGCGCAATGCTAAATTTCTGCGGAGGCACTGTACCAATTTACAGGAGAGGAATA 2160  
Db 2541 GACCGCGCAATGCTAAATTTCTGCGGAGGCACTGTACCAATTTACAGGAGAGGAATA 2600  
QY 2161 GATTTGAGGCTAACACGCTATCTTTAAAGGGAACGGTAAAGGCTGTAATATCATTTCA 2220  
Db 2601 GATTTGAGGCTAACACGCTATCTTTAAAGGGAACGGTAAAGGCTGTAATATCATTTCA 2660  
QY 2221 TCAGTGAATTAATTAACCCACAATCTTAGTGCCACAATTAACATATCTGGGAATATAACA 2280  
Db 2661 TCAGTGAATTAATTAACCCACAATCTTAGTGCCACAATTAACATATCTGGGAATATAACA 2720  
QY 2281 ATTAACCAAACTACGAGAAGAACACCTCGTATTGGCAACACCGCCATGATTCGCACTGG 2340  
Db 2721 ATTAACCAAACTACGAGAAGAACACCTCGTATTGGCAACACCGCCATGATTCGCACTGG 2780  
QY 2341 AACGTCAGTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2400  
Db 2781 AACGTCAGTGCTCTTAATCTAGAGACAGCGCAAAATTTTACCTTTTAAATACATTTCA 2840  
QY 2401 AGCAATAGCAAAAGGCTTTAAACACACAGTATAGAGCTCTCGAGGGTGAAATTAACGGC 2460  
Db 2841 AGCAATAGCAAAAGGCTTTAAACACACAGTATAGAGCTCTCGAGGGTGAAATTTAACGGC 2900  
QY 2461 GTAAATGCCAACATGTCATTCATCTCAAGAGAGGCGGAAGTTTAATTTCAAAATTAATA 2520  
Db 2901 GTAAATGCCAACATGTCATTCATCTCAAGAGAGGCGGAAGTTTAATTTCAAAATTAATA 2560  
QY 2521 CNAACGAGACATGACACACGCAACACCTTTACCAATTCGGGTTTTAGCCCAATATCACA 2580  
Db 2961 CNAACGAGACATGACACACGCAACACCTTTACCAATTCGGGTTTTAGCCCAATATCACA 3020  
QY 2581 GCCACTGGTGGGGCTCTGTTTTTTTGATATATATGCCAACCATTTCTGGCAGAGGGCT 2640  
Db 3021 GCCACTGGTGGGGCTCTGTTTTTTTGATATATATGCCAACCATTTCTGGCAGAGGGCT 3080  
QY 2641 GAGTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAAATTTTACCTTTAAATTCOCAT 2700  
Db 3081 GAGTAAAAATGAGTGAATTAATATCTCTAACGGCGCTAAATTTTACCTTTAAATTCOCAT 3140  
QY 2701 GTTCGGCGGATGACGCTTTTAAATCAACAAAGACTTAACCATATAATGCAACCAATTC 2760  
Db 3141 GTTCGGCGGATGACGCTTTTAAATCAACAAAGACTTAACCATATAATGCAACCAATTC 3200  
QY 2761 AATTTGAGGCTCAGACAGAGAGAGATGATTTTATGACGGGTACGCAACCAATGCAATC 2820  
Db 3201 AATTTGAGGCTCAGACAGAGAGAGATGATTTTATGACGGGTACGCAACCAATGCAATC 3260  
QY 2821 AATTTGAGGCTCAGACAGAGAGAGATGATTTTATGACGGGTACGCAACCAATGCAATC 2880  
Db 3261 AATTTGAGGCTCAGACAGAGAGAGATGATTTTATGACGGGTACGCAACCAATGCAATC 3320  
QY 2881 AGCAGCAGATTTACGGGGAATATTACTATCGAGAAAGAGCAAAATGTTACGCTAGAAGCC 2940  
Db 3321 AGCAGCAGATTTACGGGGAATATTACTATCGAGAAAGAGCAAAATGTTACGCTAGAAGCC 3380  
QY 2941 AATAAGCCCTAATCAGCAAAACATTAAGGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3000  
Db 3381 AATAAGCCCTAATCAGCAAAACATTAAGGGATAGAGTTATAAACTTTGGCAGCTTGCTC 3440  
QY 3001 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCAAGATATTAAAGCAATCTCACTATT 3060  
Db 3441 GTTAATGGGAGTTTAAGTTTAACTGGCGAAATGCAAGATATTAAAGCAATCTCACTATT 3500  
QY 3061 TCAGAAAGCGCACTTTAAAGCAAGACTAGAGATACCCTTAATATACCGGCAATTTT 3120

Db 3501 TCAGAAAGCGCACTTTTAAAGGAAAGACTAGAGATACCCTAAATATATACCGGCAATTTT 3560  
QY 3121 ACCAATTAATGGCACTCCGGAATTAATATACCAAGAGAGTGGTAAACTTTGGCAATGTT 3180  
Db 3561 ACCAATTAATGGCACTCCGGAATTAATATACCAAGAGAGTGGTAAACTTTGGCAATGTT 3620  
QY 3181 ACCAATTAATGGTATTAAACATTAACCACTACGCTAAACGCAACCAAGAGCATCATC 3240  
Db 3621 ACCAATTAATGGTATTAAACATTAACCACTACGCTAAACGCAACCAAGAGCATCATC 3680  
QY 3241 GCGGAGATATAATCAACAAAAAGAGAGCTTAAATATTACAGACAGTAATAATGATGCT 3300  
Db 3681 GCGGAGATATAATCAACAAAAAGAGAGCTTAAATATTACAGACAGTAATAATGATGCT 3740  
QY 3301 GAAATCCAAATTTGCGGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTCCGAT 3360  
Db 3741 GAAATCCAAATTTGCGGCAATATCTCGCAAAAGAGGCAACCTCACGATTTCTCCGAT 3800  
QY 3361 AAAATTAATTAACCAACAGATAAACAATCAAAAAGGATTTGATGGAGAGGACTTAGT 3420  
Db 3801 AAAATTAATTAACCAACAGATAAACAATCAAAAAGGATTTGATGGAGAGGACTTAGT 3860  
QY 3421 TCAGATCGCAACAGTAATGCCAACCTTAATTAACCAAGAGTAATTAATTCACAGAA 3480  
Db 3861 TCAGATCGCAACAGTAATGCCAACCTTAATTAACCAAGAGTAATTAATTCACAGAA 3920  
QY 3481 GACCTAAGTATTTTCAGGTTTCAATTAACAGAGAGTAATTAACCAAGAGTAATTAATTC 3540  
Db 3921 GACCTAAGTATTTTCAGGTTTCAATTAACAGAGAGTAATTAACCAAGAGTAATTAATTC 3980  
QY 3541 ACTATTGGCAACAGTAATGACGGTACAGCGGTCGCGGCAAGCAACAGTAATTAATTC 3600  
Db 3981 ACTATTGGCAACAGTAATGACGGTACAGCGGTCGCGGCAAGCAACAGTAATTAATTC 4040  
QY 3601 AATGTTAAAGATTTCAAAATCTCTGCTGACGGTCAACAATAGCGACAAAGT 3660  
Db 4041 AATGTTAAAGATTTCAAAATCTCTGCTGACGGTCAACAATAGCGACAAAGT 4100  
QY 3661 AAAACATCTAGCAGCAATGGCGACGTAAGCAATAGCGACAAAGT 3720  
Db 4101 AAAACATCTAGCAGCAATGGCGACGTAAGCAATAGCGACAAAGT 4160  
QY 3721 ATTACTCAAAAATGTTAGAGTAACCAAGAGATTAATTTCTCTCAAAACAGTAATTC 3780  
Db 4161 ATTACTCAAAAATGTTAGAGTAACCAAGAGATTAATTTCTCTCAAAACAGTAATTC 4220  
QY 3781 ACCGGCTCGGAAAGGTTACCAACAGAGGCTCGACCATTAACGCAACAAATGGCAAA 3840  
Db 4221 ACCGGCTCGGAAAGGTTACCAACAGAGGCTCGACCATTAACGCAACAAATGGCAAA 4280  
QY 3841 GCAAGTATTCAACCAACAGAGTGATATCAGCGGTACGATTTCCGGTTAACACCGTAAGT 3900  
Db 4281 GCAAGTATTCAACCAACAGAGTGATATCAGCGGTACGATTTCCGGTTAACACCGTAAGT 4340  
QY 3901 GTTAGCGGCACTGGTGTATTTAAACCACTAAATCCGGCTCAAAAATTTGAAGCAAAATCGGGT 3960  
Db 4341 GTTAGCGGCACTGGTGTATTTAAACCACTAAATCCGGCTCAAAAATTTGAAGCAAAATCGGGT 4400  
QY 3961 GAGGCTAATTAACAGTGCAACAGAGTACAAATTTGGCGGTACAAATTTCCGGTTAACACCGTAA 4020  
Db 4401 GAGGCTAATTAACAGTGCAACAGAGTACAAATTTGGCGGTACAAATTTCCGGTTAACACCGTAA 4460  
QY 4021 AATGTTACGCAAAACGCTGGCGATTTTAAACAGTTGGGAATGGCGAGAAATTAATGGCACA 4080  
Db 4461 AATGTTACGCAAAACGCTGGCGATTTTAAACAGTTGGGAATGGCGAGAAATTAATGGCACA 4520  
QY 4081 GAAGGAGCTCAACCTTAACCGCAACAGAGTAATACCTTGACTACTGAGCGGTTCTAGC 4140  
Db 4521 GAAGGAGCTCAACCTTAACCGCAACAGAGTAATACCTTGACTACTGAGCGGTTCTAGC 4580  
QY 4141 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGCTGCTCAGAAATGTTAGCATCGCAGGAGC 4200  
Db 4581 ATCACTTCAACTAAGGTCAGGTAGACCTCTTGCTGCTCAGAAATGTTAGCATCGCAGGAGC 4640

```

QY 4201 ATTAATGCTGCTAATGACATTAATACTACAGCACCTTAACCAAGCTGGAGGCTCG 4260
Db 4641 ATTAATGCTGCTAATGACATTAATACTACAGCACCTTAACCAAGCTGGAGGCTCG 4700
QY 4261 GATATTAAGCAACAGCGGACCTTGGTTATTAACGCAAAAGATGCTAAGCTAAATGGT 4320
Db 4701 GATATTAAGCAACAGCGGACCTTGGTTATTAACGCAAAAGATGCTAAGCTAAATGGT 4760
QY 4321 GATGCATCAGGTGATGATACAGAAAGTGAATGCACTCAACGCAAGCGGCTCGGTAGTGTG 4380
Db 4761 GATGCATCAGGTGATGATACAGAAAGTGAATGCACTCAACGCAAGCGGCTCGGTAGTGTG 4820
QY 4381 ACTGGCGCAACCTCAACGAGTGAATATCACTGGGGATTAACACAGTAATGGGTTA 4440
Db 4821 ACTGGCGCAACCTCAACGAGTGAATATCACTGGGGATTAACACAGTAATGGGTTA 4880
QY 4441 AATATCATTTGCAAGATGTAAGAACACTGTGGCTTAAGAGGCAAGAAATTTGAGGTG 4500
Db 4881 AATATCATTTGCAAGATGTAAGAACACTGTGGCTTAAGAGGCAAGAAATTTGAGGTG 4940
QY 4501 AATATATCCAGCGAGCTGTAGCAAGTGTAGCAAGTGAATTAAGCGAAACGCTCCTT 4560
Db 4941 AATATATCCAGCGAGCTGTAGCAAGTGTAGCAAGTGAATTAAGCGAAACGCTCCTT 5000
QY 4561 GAAAGATTAAGATTTATCTGATGAAGAGAGAAACATTTAGCTAAACTTGGTGAAGT 4620
Db 5001 GAAAGATTAAGATTTATCTGATGAAGAGAGAAACATTTAGCTAAACTTGGTGAAGT 5060
QY 4621 GCTGACGTTTGTGGAGCAAAATTAACAATTTACAGTCAATACACAAATGAATTTACA 4680
Db 5061 GCTGACGTTTGTGGAGCAAAATTAACAATTTACAGTCAATACACAAATGAATTTACA 5120
QY 4681 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTGTTCTCAAGTGGTAAT 4740
Db 5121 ACCAGACGCTCAAGTCAAGTGAATTTCTGAAGTGAAGCGGTGTTCTCAAGTGGTAAT 5180
QY 4741 GCGGCACAGATGATGACCAATGTTGCTGACGATGAGACGCCCTAGCTAGTAATTGACAAG 4800
Db 5181 GCGGCACAGATGATGACCAATGTTGCTGACGATGAGACGCCCTAGCTAGTAATTGACAAG 5240
QY 4801 GTAGATTTTCATCCTGCAATGAAGTCAATTTATTTTCTGATTTATTTTCTGTTGTTGTTAA 4860
Db 5241 GTAGATTTTCATCCTGCAATGAAGTCAATTTATTTTCTGATTTATTTTCTGTTGTTGTTAA 5300
QY 4861 GTTCAGTACGGGCTTTACCCATCTTTGTAAGAAATTAAGGAGAAATACAATGAATTTT 4920
Db 5301 GTTCAGTACGGGCTTTACCCATCTTTGTAAGAAATTAAGGAGAAATACAATGAATTTT 5360
QY 4921 AACAGGTTATTATTATG 4937
Db 5361 AACAGGTTATTATTATG 5377

```

## RESULT 15

```

US-08-719-641-6
; Sequence 6, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-719-641-6

```

```

Query Match 97.6%; Score 4820.2; DB 4; Length 9323;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 4864; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 TAAATATACAGAGATAATAAATAAATCAAGATTTTGTGATGACAAACAAATATACAA 60
Db 441 TAAATATACAGAGATAATAAATAAATCAAGATTTTGTGATGACAAACAAATATACAA 500
QY 61 CACCTTTTGTGAGTCTATATGCAATATTTTAAAAAATAGTATAAATCGGCATATAA 120
Db 501 CACCTTTTGTGAGTCTATATGCAATATTTTAAAAAATAGTATAAATCGGCATATAA 560
QY 121 AATGATATATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCAT 180
Db 561 AATGATATATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCAT 620
QY 181 CTTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTT 240
Db 621 CTTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTT 680
QY 241 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db 681 CACATGAATGATGAACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
QY 301 GAACCAATATGATAAGTAATTAATTTGTTCAACTAACCTTAGGAGAAAATATGAACAAG 360
Db 741 GAACCAATATGATAAGTAATTAATTTGTTCAACTAACCTTAGGAGAAAATATGAACAAG 800
QY 361 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGCTGCTGCTGAATTTGCA 420
Db 801 ATATATCGTCTCAAAATTCAGCAACCGCTGAATGCTTTGCTGCTGCTGAATTTGCA 860
QY 421 CCGGTTGTGACCATTCACACAAAAGGCTCCGCTATCTTACTATCTTTAGGTGTAAAC 480
Db 861 CCGGTTGTGACCATTCACACAAAAGGCTCCGCTATCTTACTATCTTTAGGTGTAAAC 920
QY 481 CACTTAGCGTTAAAGCCACTTTCCGCTATCTTACTATCTTTAGGTGTAAACCTATTTCCA 540
Db 921 CACTTAGCGTTAAAGCCACTTTCCGCTATCTTACTATCTTTAGGTGTAAACCTATTTCCA 980
QY 541 CAATCTGTTTTAGCAAGGGCTTTACAGGAATGGATGTAGTACACGCGACGACCTATG 600

```

||||| 981 CAATCTGTTAGCAAGCGCAATTTAACTCGACCAAAATGAAATGGTGCAGTTTATAC 1040  
QY 601 CAAGTAGATGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 660  
Db 1041 AAGAAAACAAGTAATAAACCATTATCCGCAACAGTGTGACGCTATCATTAATTTGAAA 1100  
QY 661 CAATTTACATCGACCAAAATGAATGGTGCAGTTTACAAAGAAAACAACATCCGCC 720  
Db 1101 CAATTTACATCGACCAAAATGAATGGTGCAGTTTACAAAGAAAACAACATCCGCC 1160  
QY 721 GTATTTCAACCGGTGTACATCTAACCAATCTCCCAATTTAAAGGATTTAGATCTTAAC 780  
Db 1161 GTATTTCAACCGGTGTACATCTAACCAATCTCCCAATTTAAAGGATTTAGATCTTAAC 1220  
QY 781 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATAGTAAAGAGCGCAATTTAAAC 840  
Db 1221 GGACAAGTCTTTTAAATCAACCCAAATGGTATCAATAGTAAAGAGCGCAATTTAAAC 1280  
QY 841 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 900  
Db 1281 ACTAATGGCTTTACGGCTTCTACGCTAGACATTTCTAACGAAAACATCAAGCGCGCTAAT 1340  
QY 901 TTACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAAATGTCAATCACGGTTTAAT 960  
Db 1341 TTACCTTCGAGCAACCAAGATAAAGCGCTCGCTGAAATGTCAATCACGGTTTAAT 1400  
QY 961 ACTGTCCGTAAAGAGCGGAGTGAATCTTATTTGTTGGCAAGTGAAMAGCGGGTGTG 1020  
Db 1401 ACTGTCCGTAAAGAGCGGAGTGAATCTTATTTGTTGGCAAGTGAAMAGCGGGTGTG 1460  
QY 1021 ATTAGCTAAATGGTGGCAGCATTTCTTTACTCGAGGGCAAAAATCACCATCAGCGAT 1080  
Db 1461 ATTAGCTAAATGGTGGCAGCATTTCTTTACTCGAGGGCAAAAATCACCATCAGCGAT 1520  
QY 1081 ATAAATAACCCCAACCATTTACTTACAGCATTTGCCGCGCTGAAAATGAAGCGGTCAATCTG 1140  
Db 1521 ATAAATAACCCCAACCATTTACTTACAGCATTTGCCGCGCTGAAAATGAAGCGGTCAATCTG 1580  
QY 1141 GCGCATATTTTGGCAAGGCGGTAAACATTAATGTCCGTGCTGCCACATTTTCGAAACCAA 1200  
Db 1581 GCGCATATTTTGGCAAGGCGGTAAACATTAATGTCCGTGCTGCCACATTTTCGAAACCAA 1640  
QY 1201 GGTAACTTCTGCTGATCTTAAGCAAGATTAAGCGGCAATTTGTTCTTCGCC 1260  
Db 1641 GGTAACTTCTGCTGATCTTAAGCAAGATTAAGCGGCAATTTGTTCTTCGCC 1700  
QY 1261 AAAGAGGTGAAGCGGAATTTGGCGGTGTAAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1320  
Db 1701 AAAGAGGTGAAGCGGAATTTGGCGGTGTAAATTTCCGCTCAAAATCAGCAAGCTAAAGGC 1760  
QY 1321 GSCAAGCTGATGATTACAGCGGATAAAGTTCACATTAATAAAGAGTGCAGTTATCGACCTT 1380  
Db 1761 GSCAAGCTGATGATAAAGTTCACATTAATAAAGAGTGCAGTTATCGACCTT 1820  
QY 1381 TCAGGTAAAGAGGGGGAACCTTTACCTTTGGCGGTGACGCGCGGCGGAAGGTAAAAAC 1440  
Db 1821 TCAGGTAAAGAGGGGGAACCTTTACCTTTGGCGGTGACGCGCGGCGGAAGGTAAAAAC 1880  
QY 1441 GSCATTCATTAAGCAAGAAACCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1500  
Db 1881 GSCATTCATTAAGCAAGAAACCTTTTAGAAAAAGGCTCAACCATCAATGTATCAGGC 1940  
QY 1501 AAAGAAAAGCGGAGCGCTATTCGTGGCGGATTTGGCTTAATTTAGCGGCAATAT 1560  
Db 1941 AAAGAAAAGCGGAGCGCTATTCGTGGCGGATTTGGCTTAATTTAGCGGCAATAT 2000  
QY 1561 AACGCTCAAGGTAGTGGGATATCCGTAACACCGGTGGTTTGTGGAGACATFCGGGGCAT 1620  
Db 2001 AACGCTCAAGGTAGTGGGATATCCGTAACACCGGTGGTTTGTGGAGACATFCGGGGCAT 2060  
QY 1621 TATTTATCCATTTGACAGCAATCAATTTGTTAAACAAAAGAGTGGTCTAGACCCCTGAT 1680  
|||||

Db 2061 TATTTATCCATTTGACAGCAATGCAATTTGTTAAACAAAAGAGTGGTCTAGACCCCTGAT 2120  
QY 1681 GATGTAACTTCAAGCGGAAGACCCCTTTCGCAATTAATACCGGTATAATGATGAATTC 1740  
Db 2121 GATGTAACTTCAAGCGGAAGACCCCTTTCGCAATTAATACCGGTATAATGATGAATTC 2180  
QY 1741 CCAACAGGACCGGTGAAGCAAGCGACCTTGAAGAAAATAGCGAATCAAAACAACGCTA 1800  
Db 2181 CCAACAGGACCGGTGAAGCAAGCGACCTTGAAGAAAATAGCGAATCAAAACAACGCTA 2240  
QY 1801 ACCAATACAACTATTTCAAAATTTATCTGAAGAAAAGCGCTTGAAGAAAATAGCGAATCA 1860  
Db 2241 ACCAATACAACTATTTCAAAATTTATCTGAAGAAAAGCGCTTGAAGAAAATAGCGAATCA 2300  
QY 1861 AGAAAACCTTACCCTTAAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCCAT 1920  
Db 2301 AGAAAACCTTACCCTTAAATAGCTCAATCAACATCGGAAGCAACTCCCACTTAATTTCCAT 2360  
QY 1921 AGTAAAGGTGAGGTGGCGGAGCGGTTCAGATTTGATGGAGATATTAATTTAAAGCGGA 1980  
Db 2361 AGTAAAGGTGAGGTGGCGGAGCGGTTCAGATTTGATGGAGATATTAATTTAAAGCGGA 2420  
QY 1981 AATTTAACCAATTTATCTGCGGATGGTGTGATTTTCAATAAAAATATTACGTTGATCAG 2040  
Db 2421 AATTTAACCAATTTATCTGCGGATGGTGTGATTTTCAATAAAAATATTACGTTGATCAG 2480  
QY 2041 GGTTTTTTAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAGACGC 2100  
Db 2481 GGTTTTTTAATATTACCGCGCTTCCGTAGCTTTTGAAGGTGGAATTAACAAGACGC 2540  
QY 2101 GACGCGCAATGCTAAAAATTTGCGCCAGGCACTGTAAACCATTAACAGAGAGGAAAA 2160  
Db 2541 GACGCGCAATGCTAAAAATTTGCGCCAGGCACTGTAAACCATTAACAGAGAGGAAAA 2600  
QY 2161 GATTTACAGGCTAACACAGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2220  
Db 2601 GATTTACAGGCTAACACAGCTATCTTTAAACGGAACGGGTAAAGGTCTGAATATCATTTCA 2660  
QY 2221 TCAGTGAATTAATTTAAACCCCAATCTTGTAGTGGCAATTAACATATCTGGGAATATAACA 2280  
Db 2661 TCAGTGAATTAATTTAAACCCCAATCTTGTAGTGGCAATTAACATATCTGGGAATATAACA 2720  
QY 2281 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCATGATTCGCACTGG 2340  
Db 2721 ATTAACCAACTACGAGAAAGAACACCTCGTATTGGCAACACAGCATGATTCGCACTGG 2780  
QY 2341 AACGTGAGTCTCTTAATCTAGACAGAGCGCAATTTTACCTTTATTAATACATTTCA 2400  
Db 2781 AACGTGAGTCTCTTAATCTAGACAGAGCGCAATTTTACCTTTATTAATACATTTCA 2840  
QY 2401 AGCAATAGCAAGGCTTTAAACAACACAGTATAGAGCTCTCGAGGGGTGAATTTTAAGCGC 2460  
Db 2841 AGCAATAGCAAGGCTTTAAACAACACAGTATAGAGCTCTCGAGGGGTGAATTTTAAGCGC 2900  
QY 2461 GTAAATGGCAACATGTCATTTCAATCTCAAAAGAGGAGGAAAGTAAATTTCAATTTAAAA 2520  
Db 2901 GTAAATGGCAACATGTCATTTCAATCTCAAAAGAGGAGGAAAGTAAATTTCAATTTAAAA 2960  
QY 2521 CCAAAACGAGAACATGAACACAAAGCAACCTTTACCAATTTGGTGTGAGCAATATCACA 2580  
Db 2961 CCAAAACGAGAACATGAACACAAAGCAACCTTTACCAATTTGGTGTGAGCAATATCACA 3020  
QY 2581 GCCACTGGTGGGCTCTGTTTTTGGATATATATGCCAACCATTTCTGCGAGAGGGCT 2640  
Db 3021 GCCACTGGTGGGCTCTGTTTTTGGATATATATGCCAACCATTTCTGCGAGAGGGCT 3080  
QY 2641 GAGTTAAAAATGAGTGAATTAATATCTTAACGGCGCTAAATTTACCTTAATTTCCAT 2700  
Db 3081 GAGTTAAAAATGAGTGAATTAATATCTTAACGGCGCTAAATTTACCTTAATTTCCAT 3140  
QY 2701 GTTCGCGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATAAATGCAACCAATTC 2760  
Db 3141 GTTCGCGCGATGACGCTTTTAAATCAACAAAGACTTAAACCAATAAATGCAACCAATTC 3200



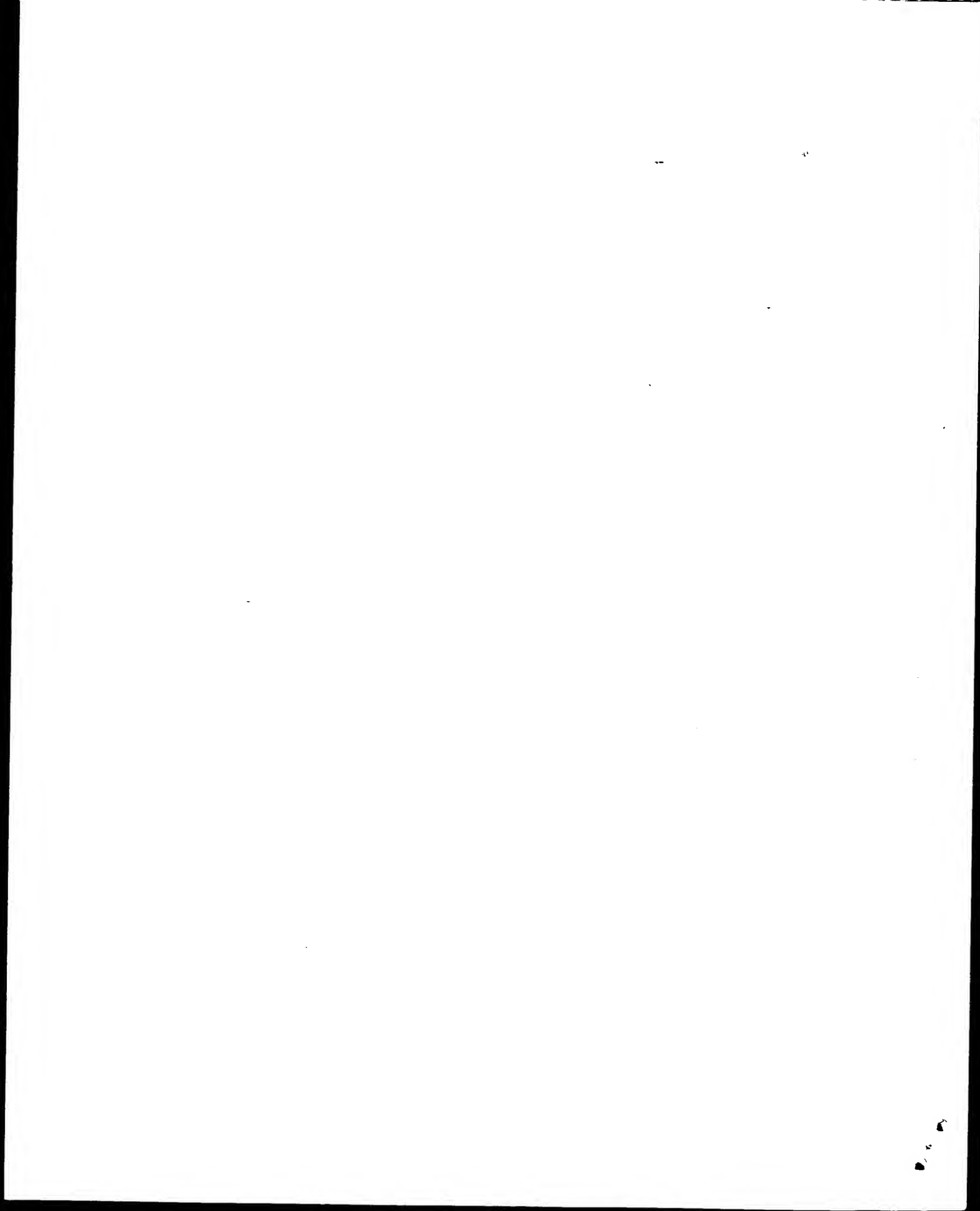
us-10-092-880-3.rni

Tue Mar 25 08:39:01 2003

Db 5361 AACAGGTTATTATG 5377  
|||||

Search completed: March 22, 2003, 05:46:10  
Job time : 258 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 20:50:22 ; Search time 6338 Seconds  
(without alignments)  
12615.512 Million cell updates/sec

Title: US-10-092-880-3

Perfect score: 4937

Sequence: 1 taaatatacaagataataaa.....tttaacaggttattattatg 4937

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_eston:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	78.2	1.6	578	17	BH000184	BH000184 2M0287B22
C 2	76.6	1.6	688	17	A2545997	A2545997 ENTEB41TR
C 3	76.2	1.5	403	17	A2654231	A2654231 1M0528K13
C 4	74.6	1.5	985	17	CNS06NAS	AL406538 T3 end of
C 5	71.2	1.4	625	17	A2605403	A2605403 1M0426P11
C 6	70.4	1.4	726	17	A2967029	A2967029 2M0237J15

7	70.4	1.4	1588	11	AK018644	Mus muscu
8	70	1.4	1101	17	CNS001T2	Drosophi
9	69.6	1.4	490	17	BH834109	BACFP6-K0
10	69.6	1.4	821	10	AV758661	AV758661
11	68.2	1.4	1101	17	CNS01F77	Drosophi
12	68.2	1.4	848	17	A2549808	ENTE262TF
13	67.6	1.4	372	17	BH478403	BOGJZ72TR
14	67.6	1.4	1063	17	CNS07A2Y	AL436064 T3 end of
15	66	1.3	342	17	AQ347002	RPC111-11
16	66	1.3	1036	17	CNS00599	AL057797 Drosophi
17	65.6	1.3	1101	17	CNS00EVL	AL069706 Drosophi
18	64.8	1.3	854	17	CNS0090S	AL052980 Drosophi
19	64.8	1.3	951	17	AZ676519	ENTE5V1TF
20	64.8	1.3	1101	17	CNS0021J	AL061936 Drosophi
21	64.6	1.3	806	17	CNS04AEE	AL281759 Tetraodon
22	64.6	1.3	1101	17	CNS0039G	AL063921 Drosophi
23	64.4	1.3	437	17	CNS030FO	AL253269 Tetraodon
24	64.2	1.3	265	13	BM274994	BM274994 PfESToa7
25	63.6	1.3	1223	17	BL2981	BL2981 T2D11-Sp6
26	63.4	1.3	388	17	CNS0026W	AL097298 Drosophi
27	63.2	1.3	613	17	AQ222590	AQ222590 RPC1-23-2
28	63.2	1.3	1001	17	CNS0135H	AL105023 Drosophi
29	63	1.3	855	17	A2532642	A2532642 ENTB49TR
30	63	1.3	944	17	CNS00A5G	AL054949 Drosophi
31	62.8	1.3	1055	14	BQ876453	BQ876453 AGENCOURT
32	62.8	1.3	1101	17	CNS000DSU	AL075280 Drosophi
33	62.6	1.3	889	17	CNS006W4	AL065999 Drosophi
34	62.4	1.3	1885	10	BE420745	BE420745 HMM002.H0
35	62	1.3	420	17	BH094146	BH094146 RPC1-24-3
36	62	1.3	974	17	CNS001TT	AL075432 Drosophi
37	62	1.3	1101	17	CNS04TO6	AL306807 Tetraodon
38	61.8	1.3	525	10	AW906883	AW906883 EST343006
39	61.8	1.3	570	12	BG860408	BG860408 1024070D0
40	61.8	1.3	604	9	AL546530	AL546530 AL546530
41	61.8	1.3	649	13	BI920072	BI920072 EST539995
42	61.8	1.3	1059	17	CNS0022B	AL097133 Drosophi
43	61.8	1.3	1300	13	BM468018	BM468018 AGENCOURT
44	61.6	1.2	417	17	CNS02AH7	AL188548 Tetraodon
45	61.6	1.2	866	17	CNS006MS	AL065764 Drosophi

#### ALIGNMENTS

RESULT 1  
BH000184/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH000184 578 bp DNA linear GSS 27-APR-2001  
2M0287B22R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
clone UUGC2M0287B22 R, DNA sequence.

BH000184

GSS.

house mouse.

Mus musculus

Eukaryota; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 578)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00





REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 625)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0426 row: P column: 11  
Seq primer: CACACAGGAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 625.  
Location/Qualifiers  
1..625  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0426P11"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

FEATURES  
source

BASE COUNT 311 a 63 c 161 g 90 t  
ORIGIN  
Query Match 1.4%; Score 71.2; DB 17; Length 625;  
Best Local Similarity 75.9%; Pred. No. 2.2e-05;  
Matches 88; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 126 TATAATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 185  
Db 324 TCT 265  
QY 186 ATCTTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 241  
Db 264 TTCTTCT 209

RESULT 6  
LOCUS A2967029 726 bp DNA linear GSS 27-APR-2001  
DEFINITION 2M0237J15R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
clone UUGC2M0237J15 R, DNA sequence.  
ACCESSION A2967029  
VERSION A2967029.1 GI:13838256

KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 726)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0237 row: J column: 15  
Seq primer: CACACAGGAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 726.  
Location/Qualifiers  
1..726  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0237J15"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, P-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

FEATURES  
source

BASE COUNT 175 a 164 c 81 g 306 t  
ORIGIN  
Query Match 1.4%; Score 70.4; DB 17; Length 726;  
Best Local Similarity 60.4%; Pred. No. 3.4e-05;  
Matches 116; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 55 TTACAAACACCTTTTTCGACGTATATGCAATATTTTAAAAAATAGTAATCCGCC 114  
Db 396 TGAGAAATCCCTTGGTGTTCACAGTTTACAAAGTTTAAATATTCATCTTCCTTT 455  
QY 115 ATATAAATGGTATATATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 174  
Db 456 CTCCTTCT 515  
QY 175 TTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC 234  
Db 516 TTTCCT 575











```

Db 949 AMTTTATTTTWTWTAAWMAATAAAATWTAWMAATAAAATAAATAAATAA 890
QY 122 ATGGTATAATCTTTTCATCTTTTAAATCTTTTCATCTTTTCATCTTTTCATC 181
  || : || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 ATWTAWAAAAATTTTWTWTATTTTWTWTATTTTWTWTATTTTWTWTATTTT 830
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 TTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 241
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 TTTTTCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 770
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 ACATGAAT 250
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AAANAMAH 761

RESULT 15
AQ347002/c
LOCUS      342 bp      DNA      linear      GSS 07-MAY-1999
DEFINITION      RPCI11-116J22-TV RPCI-11 Homo sapiens genomic clone RPCI-11-116J22,
DNA sequence.
ACCESSION      AQ347002
VERSION      AQ347002.1 GI:4171898
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 342)
AUTHORS      Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter
              J.C.
TITLE      Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
JOURNAL      Map Building
COMMENT      Unpublished (1997)
              Contact: Shaying Zhao, William Nierman, Mark Adams
              Department of Eukaryotic Genomics
              The Institute of Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: hbeetigr.org
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieterdejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
              Research Genetics (info@resgen.com). BAC end search page:
              http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
              Seq primer: T7
              Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..342
                     /organism="Homo sapiens"
                     /db_xref="GDB:7544397"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-116J22"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT      158 a      6 c      31 g      147 t
ORIGIN

Query Match      1.3%; Score 66; DB 17; Length 342;
Best Local Similarity 59.7%; Pred. No. 0.00031;
Matches 111; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 65 TTTTTCGAGTATATGCAAAATATTTTAAATAAATAGTAAATCCGCCATATAAATG 124
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TATATCATATATATATATATATATATATATATATATATATATATATATATATATC 257
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 GTATAATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTC 184
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 ATATCATATATCATATATATATATATATATATATATATATATATATATATATATAT 197
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

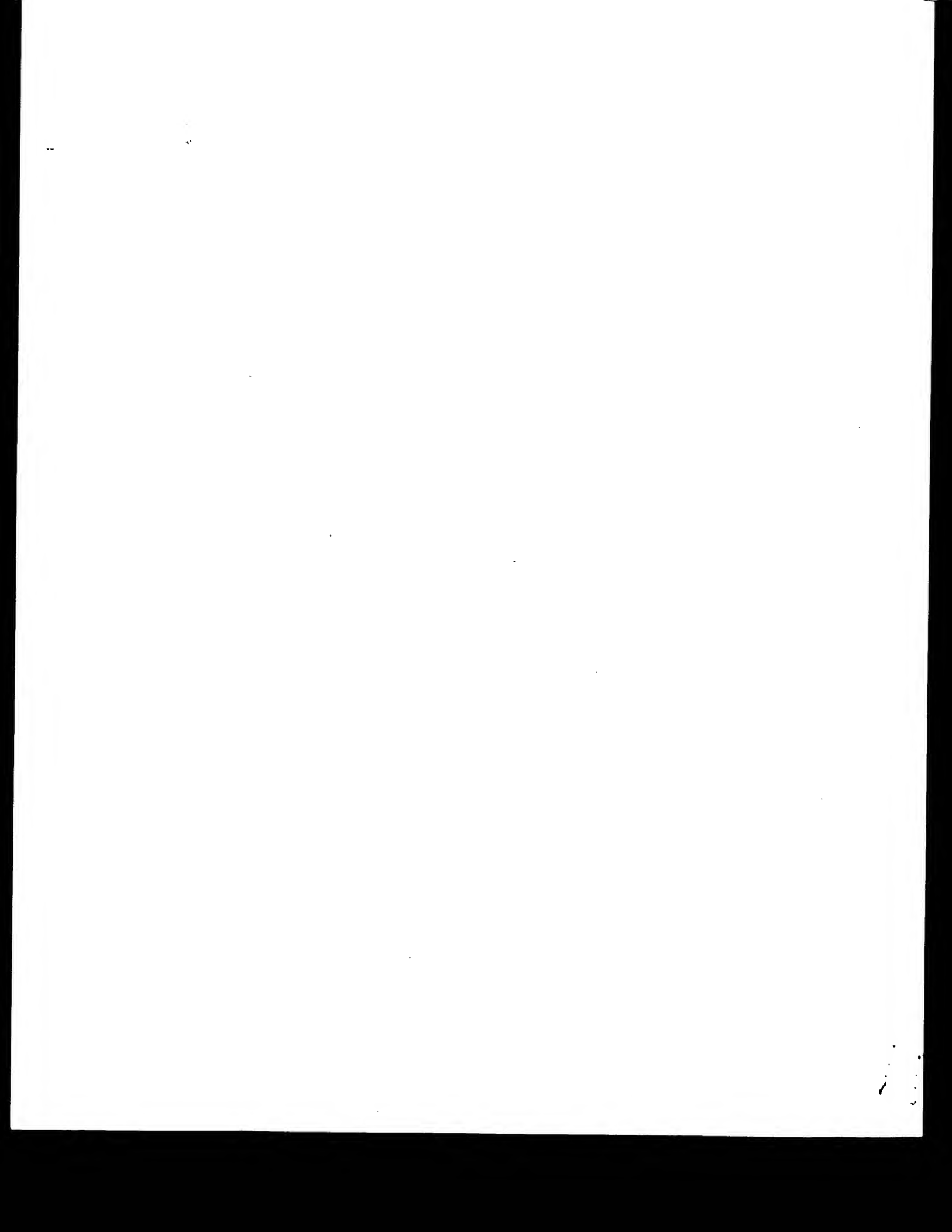
```

```

QY 185 CATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTTTCATCTTT 244
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 CATATATCATATATATATATATATATATATATATATATATATATATATATATATAT 137
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 TGAAT 250
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 TATCAT 131
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 22, 2003, 05:41:59  
Job time : 6368 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:20:59 ; Search time 46 Seconds  
(without alignments)  
4278.504 Million cell updates/sec

Title: US-10-092-880-4  
Perfect score: 7407  
Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2004.DAT:  
26: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2005.DAT:  
27: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2006.DAT:  
28: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2007.DAT:  
29: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2008.DAT:  
30: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2009.DAT:  
31: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2010.DAT:  
32: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2011.DAT:  
33: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2012.DAT:  
34: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2013.DAT:  
35: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2014.DAT:  
36: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2015.DAT:  
37: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2016.DAT:  
38: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2017.DAT:  
39: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2018.DAT:  
40: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2019.DAT:  
41: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2020.DAT:  
42: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2021.DAT:  
43: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2022.DAT:  
44: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2023.DAT:  
45: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2024.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	7389	99.8	1477	14 AAR41724	High molecular wei
2	7336	99.0	1477	18 AAW30294	Non-typeable Haemo
3	7261	98.0	1477	21 AAW01848	Haemophilus influe
4	7252	97.9	1477	15 AAR63506	Haemophilus high m
5	7193	97.1	1477	14 AAR41728	High molecular wei
6	5284	71.3	1601	18 AAW30292	Non-typeable Haemo
7	5164	69.7	1036	21 AAW01849	Haemophilus influe
8	4967	67.1	1536	14 AAR41723	High molecular wei
9	4957	66.9	1529	14 AAR41732	High molecular wei
10	4956	66.9	1536	18 AAW30293	Non-typeable Haemo

11	4943	66.7	1536	15 AAR63505	Haemophilus high m
12	4943	66.7	1536	21 AAR01846	Haemophilus influe
13	4924	66.5	1536	14 AAR41725	High molecular wei
14	4748.5	64.1	1598	18 AAR30291	Non-typeable Haemo
15	3525.5	47.6	1338	14 AAR41731	High molecular wei
16	3361.5	45.4	963	21 AAR01838	Haemophilus influe
17	3340.5	45.0	957	21 AAB01839	Haemophilus influe
18	3260	44.0	998	21 AAB01842	Haemophilus influe
19	3256.5	44.0	1079	21 AAB01836	Haemophilus influe
20	3229	43.6	992	21 AAB01843	Haemophilus influe
21	3225.5	43.5	1073	21 AAB01837	Haemophilus influe
22	3184	43.0	975	21 AAB01826	Haemophilus influe
23	3153	42.6	969	21 AAB01827	Haemophilus influe
24	2970.5	40.1	1011	21 AAB01832	Haemophilus influe
25	2939.5	39.7	1005	21 AAB01833	Haemophilus influe
26	2864	38.7	1010	21 AAB01840	Haemophilus influe
27	2833	38.2	1004	21 AAB01841	Haemophilus influe
28	2763	37.3	1095	21 AAB01847	Haemophilus influe
29	2741	37.0	1101	21 AAB01834	Haemophilus influe
30	2710	36.6	1095	21 AAB01835	Haemophilus influe
31	2678	36.2	1227	21 AAB01824	Haemophilus influe
32	2647	35.7	1221	21 AAB01825	Haemophilus influe
33	2609.5	35.2	1180	21 AAB01845	Haemophilus influe
34	2609.5	35.2	1188	21 AAB01844	Haemophilus influe
35	2592.5	35.0	1228	21 AAB01828	Haemophilus influe
36	2561.5	34.6	1222	21 AAB01830	H. influenzae stra
37	539	7.3	320	21 AAB01829	Haemophilus influe
38	517.5	7.0	1978	20 AAY27230	Amino acid sequenc
39	508	6.9	314	21 AAB01831	Haemophilus influe
40	508	6.9	1981	19 AAR42634	Protein sequence t
41	498.5	6.7	2411	21 AAR33860	Haemophilus influe
42	491	6.6	2353	17 AAR93933	Haemophilus adhesi
43	489	6.6	2123	22 AAE00701	Moraxella catarrha
44	489	6.6	3241	22 AAG66005	F. necrophorum leu
45	476	6.4	1532	20 AAY27231	Amino acid sequenc

ALIGNMENTS

RESULT 1  
AAR41724  
ID AAR41724 standard; Protein; 1477 AA.  
XX  
AC AAR41724;  
XX  
DT 26-APR-1994 (first entry)  
XX  
DE High molecular weight protein 2 (HMW2).  
XX  
KW HMW; high molecular weight protein; virus; vaccine; influenza;  
KW epitope; immunity; haemophilus influenzae.  
XX  
OS Haemophilus influenzae.  
XX  
PN WO9319090-A.  
XX  
PD 30-SEP-1993.  
XX  
PF 16-MAR-1993; 93WO-0502166.  
XX  
PR 16-MAR-1992; 92GB-0005704.  
XX  
PA (BARE/) BARENKAMP S J.  
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Barenkamp SJ;  
XX  
DR WPI: 1993-320683/40.  
XX  
N-PSDB; AAQ49507.  
XX  
PT High molecular weight surface proteins - of non-typeable  
PT haemophilus which exhibit immunogenic properties



xx  
PS  
xx  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
SQ

Claim 4; Figure 4; 100pp; English.

The isolation and purification of the high molecular weight protein enables the identification of the major protective epitopes of the protein by conventional epitope mapping. These epitopes can then be synthesised using standard techniques and incorporated into fully synthetic or recombinant vaccines.

Query Match 99.8%; Score 7389; DB 14; Length 1477;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKSEKPKARKVRLHAKPLSAMLISLGV 60  
DB 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKSEKPKARKVRLHAKPLSAMLISLGV 60

QY 61 SIPOSVLASGLQMDVHGTATMQVDGNKTIIRNSVDAIINWQFNIDQNMVQFLOENN 120  
DB 61 SIPOSVLASGLQMDVHGTATMQVDGNKTIIRNSVDAIINWQFNIDQNMVQFLOENN 120

QY 121 NSAFNVRVTSNQISQLKGLDSNGQVFLNPNGITIGKDAIINTNGFTASTLDISNENIK 180  
DB 121 NSAFNVRVTSNQISQLKGLDSNGQVFLNPNGITIGKDAIINTNGFTASTLDISNENIK 180

QY 181 ARNFTFQTKKALAEIVNHLITVVGKGSVNLIGGKVKNEGIVSVNGSSISLAGQKIT 240  
DB 181 ARNFTFQTKKALAEIVNHLITVVGKGSVNLIGGKVKNEGIVSVNGSSISLAGQKIT 240

QY 241 ISDIINPTITYSIAAPENEAENVLGDIPAKGGINVRATIRNOGKLSADSVSKDKSNIV 300  
DB 241 ISDIINPTITYSIAAPENEAENVLGDIPAKGGINVRATIRNOGKLSADSVSKDKSNIV 300

QY 301 LSAKEGEAIGGVISAQNOQAKGKLMTGDKVTLKTGAVIDLSGREGGETYVLGDERGE 360  
DB 301 LSAKEGEAIGGVISAQNOQAKGKLMTGDKVTLKTGAVIDLSGREGGETYVLGDERGE 360

QY 361 GKNGIOLAKKTSLEKSGTINVSKEKGGFAIWMGDIALIDGNINAQSGCDIAKGGFVET 420  
DB 361 GKNGIOLAKKTSLEKSGTINVSKEKGGFAIWMGDIALIDGNINAQSGCDIAKGGFVET 420

QY 421 SGHDLFIKONAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDPEPTGTGEASDPKKNSL 480  
DB 421 SGHDLFIKONAIVDAKEWLLDFDNVSIINAEDPLFNNTGINDPEPTGTGEASDPKKNSL 480

QY 481 TLTNTTISNLYKNAWTMNTASRKLTVNSNINIGSNSHLILHSGQGGGVQIDGDI 540  
DB 481 TLTNTTISNLYKNAWTMNTASRKLTVNSNINIGSNSHLILHSGQGGGVQIDGDI 540

QY 541 KGNLTIVSGGVVDVHKNTITLDQFLNITAASVAFEGGNKARDAANAKIVAOGTVITG 600  
DB 541 KGNLTIVSGGVVDVHKNTITLDQFLNITAASVAFEGGNKARDAANAKIVAOGTVITG 600

QY 601 EGKDFRANVSLNGTGLNLISSVNNLTHNLSCITINISGNITINQTRKNTSYWQTS 660  
DB 601 EGKDFRANVSLNGTGLNLISSVNNLTHNLSCITINISGNITINQTRKNTSYWQTS 660

QY 661 SHWNVSALNLETGANFTFYIKYISSNSKGLATTQYRSSAGVNFNGVNMNFKLKEGAKVNF 720  
DB 661 SHWNVSALNLETGANFTFYIKYISSNSKGLATTQYRSSAGVNFNGVNMNFKLKEGAKVNF 720

QY 721 KLPKNENMNTSKPLPREFLANITATGGSVFFDIYANHSRGAELKXSEINISGANFTL 780  
DB 721 KLPKNENMNTSKPLPREFLANITATGGSVFFDIYANHSRGAELKXSEINISGANFTL 780

QY 781 NSHVRGDDAFKINKDITINATNSFSLRQTKDDFDYGYARNAINSTYINISILGNVTLGG 840  
DB 781 NSHVRGDDAFKINKDITINATNSFSLRQTKDDFDYGYARNAINSTYINISILGNVTLGG 840

QY 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIDRVIKLSLLVNGSLSTGENADIKGN 900  
DB 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIDRVIKLSLLVNGSLSTGENADIKGN 900

DB 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIDRVIKLSLLVNGSLSTGENADIKGN 900

QY 901 LTISESATFKGTRDTLNTGNFTNNGTAEINTQGVVKLGNTDGLNITTHAKRNQR 960  
DB 901 LTISESATFKGTRDTLNTGNFTNNGTAEINTQGVVKLGNTDGLNITTHAKRNQR 960

QY 961 SLIGGDIINKKGSNLITDSNDAEIQIGGNSIQKSGENLTISDDKINITQIKKIGDGE 1020  
DB 961 SLIGGDIINKKGSNLITDSNDAEIQIGGNSIQKSGENLTISDDKINITQIKKIGDGE 1020

QY 1021 DSSSDATSNANITITKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV 1080  
DB 1021 DSSSDATSNANITITKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEAKTV 1080

QY 1081 TFNNVKDSKISADGHNVTLNSKVKTSSNGGREGSNDNDTGLTITAKNVEVKNKDI 1140  
DB 1081 TFNNVKDSKISADGHNVTLNSKVKTSSNGGREGSNDNDTGLTITAKNVEVKNKDI 1140

QY 1141 VNITASEKVTITAGTINATNGKASITTKTGDISGNTSVSATVDLTTKSGSKIEA 1200  
DB 1141 VNITASEKVTITAGTINATNGKASITTKTGDISGNTSVSATVDLTTKSGSKIEA 1200

QY 1201 KSGEANVTSATCTIGGTTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTL 1260  
DB 1201 KSGEANVTSATCTIGGTTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTL 1260

QY 1261 GSSITSTKQVDLLAQNAGSNTAGSINAAVNTLNTGTTTAVAGSDIKATSGTLVINAKDAK 1320  
DB 1261 GSSITSTKQVDLLAQNAGSNTAGSINAAVNTLNTGTTTAVAGSDIKATSGTLVINAKDAK 1320

QY 1321 LNAGSDGSTEVAVNAGSNGSVTAATSSSVNITGDLNTVNGLNITISKDGNTVRLSGKE 1380  
DB 1321 LNAGSDGSTEVAVNAGSNGSVTAATSSSVNITGDLNTVNGLNITISKDGNTVRLSGKE 1380

QY 1381 IEVKYIQPCVASVEVTEAKRVLEKVKOLDSDEERETLAKLGVSARFVPEPNTITVNTQN 1440  
DB 1381 IEVKYIQPCVASVEVTEAKRVLEKVKOLDSDEERETLAKLGVSARFVPEPNTITVNTQN 1440

QY 1441 EFTTRPSSQVIIESEKACFSSGNGARVCTNVADGQP 1477  
DB 1441 EFTTRPSSQVIIESEKACFSSGNGARVCTNVADGQP 1477

RESULT 2  
AAW30294  
ID AAW30294 standard; Protein; 1477 AA.  
XX  
AC AAW30294;  
XX  
DT 14-APR-1998 (first entry)  
XX  
DE Non-typeable Haemophilus high mol.wt. surface protein HMW2.  
XX  
KW Non-typeable Haemophilus; high molecular weight surface protein; HMW2; hmw2A gene; immunogen; vaccine; otitis media.  
XX  
OS Haemophilus influenzae strain 12.  
XX  
PH Key Location/Qualifiers  
FT Misc-difference 34 /note= "encoded by TTC"  
FT Misc-difference 35 /note= "encoded by CGC"  
FT Misc-difference 36 /note= "encoded by TAT"  
FT Misc-difference 37 /note= "encoded by GTT"  
FT Misc-difference 38 /note= "encoded by ACT"  
FT Misc-difference 39 /note= "encoded by ATC"  
FT Misc-difference 40 /note= "encoded by TTT"

FT	Misc-difference 41	/note=	"encoded by AGG"	Db	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTKGSRYVTIFRCNHLALPLSAMLISLGV	60
FT	Misc-difference 42	/note=	"encoded by TGT"	Qy	61	SIPOSVLASGLOGMDVHGHTATMQVDGKNTIIRNSVDALINNKOFNIDONAVOFLOENN	120
FT	Misc-difference 43	/note=	"encoded by AAC"	Db	61	SIPOSVLASGLOGMDVHGHTATMQVDGKNTIIRNSVDALINNKOFNIDONAVOFLOENN	120
FT	Misc-difference 44	/note=	"encoded by TAT"	Qy	121	NSAVFNRTVSNQISQLKGLDSNGQVFLINPGITIGKDAIINTNGFTASTLIDISNENIK	180
FT	Misc-difference 45	/note=	"encoded by TCC"	Db	121	NSAVFNRTVSNQISQLKGLDSNGQVFLINPGITIGKDAIINTNGFTASTLIDISNENIK	180
FT	Misc-difference 46	/note=	"encoded by GAC"	Qy	181	ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSGISILLAGOKIT	240
FT	Misc-difference 47	/note=	"encoded by GAC"	Db	181	ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSGISILLAGOKIT	240
FT	Misc-difference 48	/note=	"encoded by AGC"	Qy	241	ISDIINPTITYSTAAPENEAVNLGDIKFAKGGNINVRATIRNOGKLSADSVSKDKSGNIV	300
FT	Misc-difference 49	/note=	"encoded by AAA"	Db	241	ISDIINPTITYSTAAPENEAVNLGDIKFAKGGNINVRATIRNOGKLSADSVSKDKSGNIV	300
FT	Misc-difference 50	/note=	"encoded by ACA"	Qy	301	LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGETYLGGERGE	360
FT	Misc-difference 51	/note=	"encoded by CCT"	Db	301	LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGETYLGGERGE	360
FT	Misc-difference 52	/note=	"encoded by GAT"	Qy	361	GKNGIOLAKKTSLEKGSTINVSQKEKGGAIVWGDIALIDGNINAOGSGDIAGTGGFVET	420
FT	Misc-difference 53	/note=	"encoded by ACA"	Db	361	GKNGIOLAKKTSLEKGSTINVSQKEKGGAIVWGDIALIDGNINAOGSGDIAGTGGFVET	420
FT	Misc-difference 54	/note=	"encoded by GAA"	Qy	421	SGHDLFIKDNAIVDAKEWLLDFDNVINAEDPLFNNTGINDPEPTGTGEASDPKKNSELK	480
FT	Misc-difference 55	/note=	"encoded by CGC"	Db	421	SGHDLFIKDNAIVDAKEWLLDFDNVINAEDPLFNNTGINDPEPTGTGEASDPKKNSELK	480
XX	WO9736914-A1.			Db	481	TTLTNTTISNYLKNWMTNITASRKLTVNSSINIGNSHLILHSGKQGGVQIDGDITS	540
PN	09-OCT-1997.			Db	481	TTLTNTTISNYLKNWMTNITASRKLTVNSSINIGNSHLILHSGKQGGVQIDGDITS	540
XX	01-APR-1997; 97WO-US04707.			Qy	541	KGGNLTIIYSGGWVDVHKNTITLDQGFNLITAAASVAFEGGNKKAADAANAKIVAQCTVTITG	600
XX	01-APR-1996; 96US-0617697.			Db	541	KGGNLTIIYSGGWVDVHKNTITLDQGFNLITAAASVAFEGGNKKAADAANAKIVAQCTVTITG	600
PA	(BARE/) BARENKAMP S J.			Qy	601	EGKDFRANNVSLGTGKGLNIISVNNLTHNLSTGINSITINSGNITINOTTRKNTSYWQTSMD	660
PI	Barenkamp SJ;			Db	601	EGKDFRANNVSLGTGKGLNIISVNNLTHNLSTGINSITINSGNITINOTTRKNTSYWQTSMD	660
XX	WPI; 1997-503038/46.			Qy	661	SHWVNSALNLETGANFTFIKYISSNSKGLTIQYRSSAGVNFVNGVNGMSENLKGAQVNF	720
DR	N-PSDB; AAT90995.			Db	661	SHWVNSALNLETGANFTFIKYISSNSKGLTIQYRSSAGVNFVNGVNGMSENLKGAQVNF	720
PT	High molecular weight proteins of non-typeable Haemophilus			Qy	721	KLKPNENMNTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMSSEINISNGANFTL	780
PT	influenzae - useful for vaccine production			Db	721	KLKPNENMNTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMSSEINISNGANFTL	780
PS	Claim 7; Page 73-78; 183pp; English.			Qy	781	NSHVRGDDAFKINKDLTINATNSNFSLRQTKDDFYDGYARNAINSTYNTSILGNNVTLAGG	840
XX	This protein comprises the high molecular weight surface protein			Db	781	NSHVRGDDAFKINKDLTINATNSNFSLRQTKDDFYDGYARNAINSTYNTSILGNNVTLAGG	840
CC	HMW2 (123 kDa) of non-typeable Haemophilus influenzae strain 12 that			Qy	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDVIRKIGSLVNGSLSLTGENADIKGN	900
CC	has the immunological ability to protect against disease caused by a			Db	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDVIRKIGSLVNGSLSLTGENADIKGN	900
CC	non-typeable Haemophilus strain and is characterised by at least			Qy	901	LTISATFPGKTRDTLNTITGNTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGT	960
CC	one surface-exposed B-cell epitope that is recognised by monoclonal			Db	901	LTISATFPGKTRDTLNTITGNTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGT	960
CC	antibody AD6. The HMW2 amino acid sequence was deduced from the			Qy	961	SIIGDDIINKKGLSLNITDSNDAEIOIGGNIISQEGNLTITSSDKINTTKQITIKKIDGE	1020
CC	hmw2 gene sequence (see AAT90995 and AAT90997). The expressed protein			Db	961	SIIGDDIINKKGLSLNITDSNDAEIOIGGNIISQEGNLTITSSDKINTTKQITIKKIDGE	1020
CC	is truncated, starting at residue 442 of the full-length gene			Qy	1021	DSSSDATSNANLTITKELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080
CC	product. HMW1 (see AAW30293), HMW3 (see AAW30291) and HMW4 (see			Db	1021	DSSSDATSNANLTITKELKLTEDLSISGFNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080
CC	AAW30292) have also been identified. A conjugate comprising HMW2			Qy	1081	TFNNVKSOKISADGHNVTLSKVKVTSNGGREGSNDNDTGLTITAKNWNKIDITSLKT	1140
CC	linked to an antigen, haptan or polysaccharide, and a synthetic			Db	1081	TFNNVKSOKISADGHNVTLSKVKVTSNGGREGSNDNDTGLTITAKNWNKIDITSLKT	1140
CC	peptide of 6-150 amino acids corresponding to at least protective			Qy			
CC	epitope of HMW2 are also claimed. HMW proteins, conjugates and			Db			
CC	peptides can be used in vaccines, as immunogens for preparation of			Qy			
CC	antibodies and as antigens for detection of these antibodies.			Db			
XX	Sequence 1477 AA;			Qy			
XX	Query Match			Db			
XX	Best Local Similarity 99.0%; Score 7336; DB 18; Length 1477;			Qy			
XX	Matches 1465; Conservative 1; Mismatches 11; Indels 0; Gaps 0;			Db			
QY	1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKGSRYVTIFRCNHLALPLSAMLISLGV			QY			

QY 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISOTISGNTVSVSATVDLTTKSGSKIEA 1200  
 Db 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISOTISGNTVSVSATVDLTTKSGSKIEA 1200  
 QY 1201 KSGEANTVATGTTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTANTLTTEA 1260  
 Db 1201 KSGEANTVATGTTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTANTLTTEA 1260  
 QY 1261 GSSITSTKGVLLAQSISAGSINAANVTNTGTLTFTVAGSDIKATGTLVINAKDAK 1320  
 Db 1261 GSSITSTKGVLLAQSISAGSINAANVTNTGTLTFTVAGSDIKATGTLVINAKDAK 1320  
 QY 1321 LNGDASDSTEYNNAVNSGSGSVTAATSSSVNITGDLTVNGNLISKGRNVTLRGKE 1380  
 Db 1321 LNGDASDSTEYNNAVNSGSGSVTAATSSSVNITGDLTVNGNLISKGRNVTLRGKE 1380  
 QY 1381 IEVKYIQPGVASVEEIVIAKRVLEKVKDLSDEERTLAKLGVSAVRFPNNTITVNTQN 1440  
 Db 1381 IEVKYIQPGVASVEEIVIAKRVLEKVKDLSDEERTLAKLGVSAVRFPNNTITVNTQN 1440  
 QY 1441 EFTTRPSSQVIIEGKACFSGNGARVCTNVADGQP 1477  
 Db 1441 EFTTRPSSQVIIEGKACFSGNGARVCTNVADGQP 1477

## RESULT 3

AA01848  
 ID AA01848 standard; Protein; 1477 AA.  
 AC AA01848;  
 DT 11-SEP-2000 (first entry)  
 DE Haemophilus influenzae strain 12 HmW2A protein, SEQ ID NO:71.  
 XX  
 KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
 KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;  
 KW recombinant production; Escherichia coli; antibacterial; vaccine;  
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
 KW detection; diagnosis.  
 XX  
 OS Haemophilus influenzae strain 12.  
 XX  
 PN WO200020609-A2.  
 XX  
 PD 13-APR-2000.  
 XX  
 PF 07-OCT-1999; 99WO-CA00938.  
 XX  
 PR 07-OCT-1998; 98US-0167568.  
 PR 08-DEC-1998; 98US-0206942.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore SM, Yang Y, Klein MH;  
 XX  
 XX WPI: 2000-303789/26.  
 DR N-PSDB; AA52197.  
 XX

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

Example 16; Fig 29A-N; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene  
 CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,

CC and the hmwB and hmwC genes encode the structural HMW proteins  
 CC responsible for post-translational processing and secretion of the HMW  
 CC proteins. The modified hmwABC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMW. The invention also discloses hmwA genes (AA52175-AA52198)  
 CC and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae  
 CC strains J9yc, K1, K21, LDCD2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae and  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMW protein from a non-typeable strain of  
 CC H. influenzae.

XX  
 SQ Sequence 1477 AA;

Query Match 98.0%; Score 7261; DB 21; Length 1477;  
 Best Local Similarity 98.4%; Pred. No. 0;  
 Matches 1454; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSEKPARMKVHLALPLSALLSLGVT 60  
 Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSEKPARMKVHLALPLSALLSLGVT 60

QY 61 9IPQSVLASGLQGMVYVHGVTATMQVDGKNTIIRNSVDAILNMKQFNIDQNMVQFLOENN 120  
 Db 61 9IPQSVLASGLQGMVYVHGVTATMQVDGKNTIIRNSVDAILNMKQFNIDQNMVQFLOENN 120

QY 121 NSAVFNRTVSNQISOLKGLDSNGOVFLINPNTIGTKDALTINTNGFTASTLDSNENIK 180  
 Db 121 NSAVFNRTVSNQISOLKGLDSNGOVFLINPNTIGTKDALTINTNGFTASTLDSNENIK 180

QY 181 ARNFTFEOTKOKALAEIVNHGLITVKGSGVNLGKVKNEGIVSNGSGISLLAGOKIT 240  
 Db 181 ARNFTFEOTKOKALAEIVNHGLITVKGSGVNLGKVKNEGIVSNGSGISLLAGOKIT 240

QY 241 ISDIINPTITYTIAAPENEAVALNGDIFAKGGINVRAATIRNOGKLSDSVSKDKSGNIV 300  
 Db 241 ISDIINPTITYTIAAPENEAVALNGDIFAKGGINVRAATIRNOGKLSDSVSKDKSGNIV 300

QY 301 LSAKGEAEIGGVISAQNOQAKGGLMTGDKVTLKTGAVIDLSGKEGGETYLGGERGE 360  
 Db 301 LSAKGEAEIGGVISAQNOQAKGGLMTGDKVTLKTGAVIDLSGKEGGETYLGGERGE 360

QY 361 GKNGIOLAKKTSLEKSGSTINVSQKKGFAIWMGDIALDGNINAQSGSDIAKTGGFVET 420  
 Db 361 GKNGIOLAKKTSLEKSGSTINVSQKKGFAIWMGDIALDGNINAQSGSDIAKTGGFVET 420

QY 421 SGHDLFIKDNAIVDAKEWLLDFDNVSVINAEPLFNNTGINDPEPTGCEASDPKKNSELK 480  
 Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSVINAEPLFNNTGINDPEPTGCEASDPKKNSELK 480

QY 481 TLTNTTISNLYKNAMTMITASRKLTVNSSINIGNSHLILHSKGGGQVQIDGDIIS 540  
 Db 481 TLTNTTISNLYKNAMTMITASRKLTVNSSINIGNSHLILHSKGGGQVQIDGDIIS 540

QY 541 KGNLTIIYSGGWDVHKNTITLDQGFNLITAAVAFEGGNNKARDAANAIAVQGTVTITG 600  
 Db 541 KGNLTIIYSGGWDVHKNTITLDQGFNLITAAVAFEGGNNKARDAANAIAVQGTVTITG 600

QY 601 ECKDFRANNVSLNGTKGLNISSVNNLTNHLSGTINISGNITINOTFRKNTSYWQTSID 660  
 Db 601 ECKDFRANNVSLNGTKGLNISSVNNLTNHLSGTINISGNITINOTFRKNTSYWQTSID 660

QY 661 SHWNYSALNLETGANFTFKYISSNSKGLTTOYRSSAGVNFNGVNGNNSFLNKEGAKVNF 720  
 Db 661 SHWNYSALNLETGANFTFKYISSNSKGLTTOYRSSAGVNFNGVNGNNSFLNKEGAKVNF 720

QY 721 KLPENNMNTSKPLPIRFLANITATGGSVFFDIYANHSRGAEKLMSEINISNGANFTL 780  
 DB 721 KLPENNMNTSKPLPIRFLANITATGGSVFFDIYANHSRGAEKLMSEINISNGANFTL 780  
 QY 781 NSHVGGDAFKINKDLINATNSFSLRQTKDDFYDGYARNAINSTYNSILGNNVTLLG 840  
 DB 781 NSHVGGDAFKINKDLINATNSFSLRQTKDDFYDGYARNAINSTYNSILGNNVTLLG 840  
 QY 841 QNSSSSITGNTIEKAANVTLEANNAPNOQNIRDRVTKLGSLLVNGSLTGENADTKGN 900  
 DB 841 QNSSSSITGNTIEKAANVTLEANNAPNOQNIRDRVTKLGSLLVNGSLTGENADTKGN 900  
 QY 901 LTISEATFKGTRDTLNTITGNFTNGTAEINITOGVVKLGNTVNDGDLNITTHAKRNOR 960  
 DB 901 LTISEATFKGTRDTLNTITGNFTNGTAEINITOGVVKLGNTVNDGDLNITTHAKRNOR 960  
 QY 961 SIIGDIIKKGSLNITDSNNDAEIQGGNISQKEGNTLTSSDKINIKQITTKKGDGE 1020  
 DB 961 SIIGDIIKKGSLNITDSNNDAEIQGGNISQKEGNTLTSSDKINIKQITTKKGDGE 1020  
 QY 1021 DSSSDATSNANITKTTELKLTEDLSISGFNKAETITAKDGRDLTIGNSDNGSCAEAKTV 1080  
 DB 1021 DSSSDATSNANITKTTELKLTEDLSISGFNKAETITAKDGRDLTIGNSDNGSCAEAKTV 1080  
 QY 1081 TFNVVKDSKISADGHNVTLNSKVTTSSNGGREGSNDTGLTITAKNVEVNRKIDTSLT 1140  
 DB 1081 TFNVVKDSKISADGHNVTLNSKVTTSSNGGREGSNDTGLTITAKNVEVNRKIDTSLT 1140  
 QY 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200  
 DB 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200  
 QY 1201 KSGEANVTSAITGTTGTTGTTGNVTNVNATNAGDLTVGNGAEINATEGAATLTATGNTLTEA 1260  
 DB 1201 KSGEANVTSAITGTTGTTGTTGTTGNVTNVNATNAGDLTVGNGAEINATEGAATLTATGNTLTEA 1260  
 QY 1261 GSSITSTKGQVDDLLAONGSIAGSINAAVNTLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320  
 DB 1261 GSSITSTKGQVDDLLAONGSIAGSINAAVNTLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320  
 QY 1321 LMGDASGDSPTENVANASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTVRLRGKE 1380  
 DB 1321 LMGDASGDSPTENVANASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTVRLRGKE 1380  
 QY 1381 IEVKYIQPCVASVEVIEAKRVLEKVDLSDEERETLAKGVSAVRFEVPNNTITVNTQN 1440  
 DB 1381 IEVKYIQPCVASVEVIEAKRVLEKVDLSDEERETLAKGVSAVRFEVPNNTITVNTQN 1440  
 QY 1441 EFTTRPPSQVVISSEKACFSSGNGARVCTNVADGQP 1477  
 DB 1441 EFTTRPPSQVVISSEKACFSSGNGARVCTNVADGQP 1477  
 RESULT 4  
 AAR63506  
 ID AAR63506 standard; Protein; 1477 AA.  
 XX  
 AC AAR63506;  
 XX  
 DT 25-JUN-1995 (first entry)  
 XX  
 DE Haemophilus high molecular weight protein HMW2.  
 XX  
 KW High molecular weight protein; HMW2; protective vaccine; otitis;  
 KW sinusitis; bronchitis; Hib; ss.  
 XX  
 OS Haemophilus.  
 XX  
 PN W09421290-A.  
 XX  
 PD 29-SEP-1994.  
 XX

PF 15-MAR-1994; 94WO-US02550.  
 XX  
 PR 16-MAR-1993; 93US-0038682.  
 XX  
 PA (BARE/) BARENKAMP S J.  
 PA (SGEM/) ST GEME J W.  
 XX  
 PI Barenkamp SJ, St GEME JW;  
 XX  
 DR WPI; 1994-316665/39.  
 DR Q-PSDB; Q72294.  
 XX  
 PT New immunogenic high mol. wt. proteins of non typeable  
 PT Haemophilus - useful in protective vaccines  
 XX  
 PS Claim 3; Page 36; 127pp; English.  
 XX  
 CC The HMW2 protein encoded by this sequence is useful in a vaccine to  
 CC protect against disease caused by non-typeable Haemophilus which are  
 CC not controlled by H. influenzae type b (Hib) vaccines. The encoded  
 CC protein can also be used as a carrier for protective Hib  
 CC polysaccharide (in a conjugate vaccine against meningitis) or for  
 CC other antigens, haptens, etc.  
 CC  
 SQ Sequence 1477 AA;  
 Query Match 97.9%; Score 7252; DB 15; Length 1477;  
 Best Local Similarity 98.3%; Pred. No. 0;  
 Matches 1452; Conservative 4; Mismatches 21; Indels 0; Gaps 0;  
 QY 1 MNKIYRLKFSKRNLNVLAVSELARGCDHSTEGSEKPARMKVHRLKLPKLSAMLLSLGVT 60  
 DB 1 MNKIYRLKFSKRNLNVLAVSELARGCDHSTEGSEKPARMKVHRLKLPKLSAMLLSLGVT 60  
 QY 61 SIPOSVLASLOGMDVHVHGTATMOVDGNKTIIRNSVDALINMKQFNIDONEMVQFLOENN 120  
 DB 61 SIPOSVLASLOGMDVHVHGTATMOVDGNKTIIRNSVDALINMKQFNIDONEMVQFLOENN 120  
 QY 121 NSAVFNRTVSNQISQLKGLDSNGQVFLNPNGITIGKDAIINTNGFTASTLDSINENIK 180  
 DB 121 NSAVFNRTVSNQISQLKGLDSNGQVFLNPNGITIGKDAIINTNGFTASTLDSINENIK 180  
 QY 181 ARNFTTEOTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISYNGGSISLLAGOKIT 240  
 DB 181 ARNFTTEOTKDKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISYNGGSISLLAGOKIT 240  
 QY 241 ISDIINPTITYSTAAPENEAENGLDIFAKGGINVRAATIRNOGKLSADSVSKDKSGNIV 300  
 DB 241 ISDIINPTITYSTAAPENEAENGLDIFAKGGINVRAATIRNOGKLSADSVSKDKSGNIV 300  
 QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGBERGE 360  
 DB 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGREGGETYLGGBERGE 360  
 QY 361 GKNGIOLAKKTSLEKSTINVSQKEKGGFAIVWGDIALIDGNINAQSGDIAKTGGFVET 420  
 DB 361 GKNGIOLAKKTSLEKSTINVSQKEKGGFAIVWGDIALIDGNINAQSGDIAKTGGFVET 420  
 QY 421 SGHDLFTKONAIYDAKEWLLDFDNVSINAEDPLFNNTGTINDEFTPTGTGEASDPKKNSELK 480  
 DB 421 SGHDLFTKONAIYDAKEWLLDFDNVSINAEDPLFNNTGTINDEFTPTGTGEASDPKKNSELK 480  
 QY 481 TLTNTTTSNLYKNAMTMTNITASRKLTIVNSSINIGNSHLLILHSKGORGGGVQIDGDTITS 540  
 DB 481 TLTNTTTSNLYKNAMTMTNITASRKLTIVNSSINIGNSHLLILHSKGORGGGVQIDGDTITS 540  
 QY 541 KGNLTIYSGGVVDVHKNTITLDQGFNLITAAVSFAFEGGNKKAADAANAKIVAGTWTITC 600  
 DB 541 KGNLTIYSGGVVDVHKNTITLDQGFNLITAAVSFAFEGGNKKAADAANAKIVAGTWTITC 600  
 QY 601 EGKDFRANVSLNGTGKGLNIISVNNLTHNLSGTINISGNITINOTTRKNTSYWOTSHD 660  
 DB 601 EGKDFRANVSLNGTGKGLNIISVNNLTHNLSGTINISGNITINOTTRKNTSYWOTSHD 660

QY 661 SHNVSALETCANFETIKYISSNKGTLTQYRSSAGVNFNGVNGMSFNLKEGAKVNF 720  
 DB 661 SHNVSALETCANFETIKYISSNKGTLTQYRSSAGVNFNGVNGMSFNLKEGAKVNF 720  
 QY 721 KLPKNMNTSKPLPIREFLANITATGGSVFFFDIYANHSRGAELKMSSEINISNGANFTL 780  
 DB 721 KLPKNMNTSKPLPIREFLANITATGGSVFFFDIYANHSRGAELKMSSEINISNGANFTL 780  
 QY 781 NSHVGGDAFKINKDLTINATNSFSLRQTKDDFYDCYARNAINSTYINISILGNNVTLLG 840  
 DB 781 NSHVGGDAFKINKDLTINATNSFSLRQTKDDFYDCYARNAINSTYINISILGNNVTLLG 840  
 QY 841 QNSSSITGNTITTEKAANYTLLEANNAPNOONIRDRVILKGLSLLVNGSLSTGENADIKGN 900  
 DB 841 QNSSSITGNTITTEKAANYTLLEANNAPNOONIRDRVILKGLSLLVNGSLSTGENADIKGN 900  
 QY 901 LTISESATFKGTRDTLNTITGNTNNGTAEINITQGVVKGVLNVTNDGDLNITTHAKRNOR 960  
 DB 901 LTISESATFKGTRDTLNTITGNTNNGTAEINITQGVVKGVLNVTNDGDLNITTHAKRNOR 960  
 QY 961 SIIGGDIINKKGLSLNTDSNDAEIQIGGNISQKEGNTLISDDKINITKQITIKKGIDGE 1020  
 DB 961 SIIGGDIINKKGLSLNTDSNDAEIQIGGNISQKEGNTLISDDKINITKQITIKKGIDGE 1020  
 QY 1021 DSSSDATSNANLTIKTTELKLTEDLSISGFNKAETITAKDGRDLTIGNSDNGSGAEAKTV 1080  
 DB 1021 DSSSDATSNANLTIKTTELKLTEDLSISGFNKAETITAKDGRDLTIGNSDNGSGAEAKTV 1080  
 QY 1081 TFNNVKDKISADGHNVTLNSKVKTSSNGGRESNDNDTGLTITAKNVEYNKDITSLKT 1140  
 DB 1081 TFNNVKDKISADGHNVTLNSKVKTSSNGGRESNDNDTGLTITAKNVEYNKDITSLKT 1140  
 QY 1141 VNITASEKVTTTGTATINATNGKASITTKTGDISTISGNTVSVSATVDLTITKSGSKIEA 1200  
 DB 1141 VNITASEKVTTTGTATINATNGKASITTKTGDISTISGNTVSVSATVDLTITKSGSKIEA 1200  
 QY 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260  
 DB 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260  
 QY 1261 GSSITSTKGVDLLAQNCSIASGSIAGSINAAVNTLTGTLTITVAGSDIKATSGTLVINAKDAK 1320  
 DB 1261 GSSITSTKGVDLLAQNCSIASGSIAGSINAAVNTLTGTLTITVAGSDIKATSGTLVINAKDAK 1320  
 QY 1321 LNGDASGDSSTEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNITISKDRNTVRLRGE 1380  
 DB 1321 LNGDASGDSSTEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNITISKDRNTVRLRGE 1380  
 QY 1381 IEVKYIQGVASVEEVIEAKRVLEKVDLSDEERETLAKLGVSAVRFEPPNNTITVNTQN 1440  
 DB 1381 IEVKYIQGVASVEEVIEAKRVLEKVDLSDEERETLAKLGVSAVRFEPPNNTITVNTQN 1440  
 QY 1441 EFTTRPSSQVLIISGKACFSSGNGARVCTNVADGQP 1477  
 DB 1441 EFTTRPSSQVLIISGKACFSSGNGARVCTNVADGQP 1477

## RESULT 5

AAR41728

ID AAR41728 standard; Protein; 1477 AA.

XX AC

XX AC

XX DT

XX DT

XX DE

XX High molecular weight protein 2 (HMW2).

XX KW

XX KW

XX OS

XX OS

XX

XX

PN W09319090-A.  
 XX 30-SEP-1993.  
 XX 16-MAR-1993; 93WO-US02166.  
 XX 16-MAR-1992; 92GB-0005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM ) INERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp SJ;

XX WPI; 1993-320683/40.

XX N-PSDB; AAQ49509.

XX High molecular weight surface proteins - of non-typeable

XX haemophilus which exhibit immunogenic properties

XX Claim 4; Figure 4; 100pp; English.

XX The isolation and purification of the high molecular weight protein

XX enables the identification of the major protective epitopes of the

XX protein by conventional epitope mapping. These epitopes can then be

XX synthesised using standard techniques and incorporated into fully

XX synthetic or recombinant vaccines. This sequence is claimed to be

XX the same as that given in AAR41724 (High molecular weight protein 2)

XX although it does differ slightly.

XX SQ Sequence 1477 AA;

XX Query Match 97.1%; Score 7193; DB 14; Length 1477;

XX Best Local Similarity 97.8%; Pred. No. 0;

XX Matches 1444; Conservative 5; Mismatches 28; Indels 0; Gaps 0;

QY 1 MNKIVRLKFSKRLNALVAVSELARGCDHSTEGSEKPAKMKVHRHLAKPLSAMLLSLGVT 60

DB 1 MNKIVRLKFSKRLNALVAVSELARGCDHSTEGSEKPAKMKVHRHLAKPLSAMLLSLGVT 60

QY 61 SIPOSVLASGLQMDVHVHGTATMVDGNKTIIRNSVDALINNKQFNDQNMVQFLOENN 120

DB 61 SIPOSVLASGLQMDVHVHGTATMVDGNKTIIRNSVDALINNKQFNDQNMVQFLOENN 120

QY 121 NSAVENRVTNQISOQLKGLDSNGQVFLNPNGITIGKDAIINTNGFTASTLDISNENIK 180

DB 121 NSAVENRVTNQISOQLKGLDSNGQVFLNPNGITIGKDAIINTNGFTASTLDISNENIK 180

QY 181 ARNFTFEQTKDALKAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSGISLLAGOKIT 240

DB 181 ARNFTFEQTKDALKAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSGISLLAGOKIT 240

QY 241 ISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSVSKDSGNIV 300

DB 241 ISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSVSKDSGNIV 300

QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSCKEGGETYVLGDERGE 360

DB 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSCKEGGETYVLGDERGE 360

QY 361 GKNIGIOLAKTSLBKSTINVSKEKGFIAVWGDIALIDGNINAAQSGDIAKTGGFVET 420

DB 361 GKNIGIOLAKTSLBKSTINVSKEKGFIAVWGDIALIDGNINAAQSGDIAKTGGFVET 420

QY 421 SGHDLFIKDNAIVDAKEWLLDFDQVSVNAEDPLFNNTGINDFPTGTGEADSPKKNSEK 480

DB 421 SGHDLFIKDNAIVDAKEWLLDFDQVSVNAEDPLFNNTGINDFPTGTGEADSPKKNSEK 480

QY 481 TTLTNTTISYLNKNAWTNITASKRLTVNSINIGSNHLSKQKRGVQIDGDTITS 540

DB 481 TTLTNTTISYLNKNAWTNITASKRLTVNSINIGSNHLSKQKRGVQIDGDTITS 540

QY 541 KGGNLTISYGGVVDVHKNTLTDQGLNITATASVAFEGGNKARDANAKIIVAGTIVTITG 600

DB 541 KGGNLTISYGGVVDVHKNTLTDQGLNITATASVAFEGGNKARDANAKIIVAGTIVTITG 600

```

Db 541 KGNLTYSGGWVDVHKNTLDGFLNITTAASVAFEGGNKARDAANAQIVTITG 600
Qy 601 EGKDFRANVSLNGTGKGLNLISSVNNLTHNLSTGNTINISGNITNTOTTRKNTSYQWTSID 660
Db 601 EGKDFRANVSLNGTGKGLNLISSVNNLTHNLSTGNTINISGNITNTOTTRKNTSYQWTSID 660
Qy 661 SHNVSALENETGANFTFKYISSNKGLETFQYRSSAGVNFNGVNMFSNFKEGAKVNF 720
Db 661 SHNVSALENETGANFTFKYISSNKGLETFQYRSSAGVNFNGVNMFSNFKEGAKVNF 720
Qy 721 KLPENNMNTSKPLPIRFLANITATGGSVFFDIYAHNSGRGAELKMSKSEINISNGANFTL 780
Db 721 KLPENNMNTSKPLPIRFLANITATGGSVFFDIYAHNSGRGAELKMSKSEINISNGANFTL 780
Qy 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNTISILGNNVTLLG 840
Db 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNTISILGNNVTLLG 840
Qy 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIIRDVVKLGSLLVNGSLTGTENADIKGN 900
Db 841 QNSSSITGNITTEKAANVTLEANNAPNOQNIIRDVVKLGSLLVNGSLTGTENADIKGN 900
Qy 901 LTISEATFKGKPTDNLITNGTNTNGTAEINITQGVVKLGNTVNDGDLNITTHAKRNOR 960
Db 901 LTISEATFKGKPTDNLITNGTNTNGTAEINITQGVVKLGNTVNDGDLNITTHAKRNOR 960
Qy 961 SIIGDDIINKGSLNITDSNDAEIQGGNISQEGNLTSSDKINITKOITIKKGIDGE 1020
Db 961 SIIGDDIINKGSLNITDSNDAEIQGGNISQEGNLTSSDKINITKOITIKKGIDGE 1020
Qy 1021 DSSSDATSNANLTKTKELKLTEDLSISGFKAEITAKDGRDLTIGNSDNGSGAEAKTV 1080
Db 1021 DSSSDATSNANLTKTKELKLTEDLSISGFKAEITAKDGRDLTIGNSDNGSGAEAKTV 1080
Qy 1081 TFNVKDSKISADGHNVTLNSKVTSSNGSGRESNNDTGLTITAKNVEVNDKITSKLT 1140
Db 1081 TFNVKDSKISADGHNVTLNSKVTSSNGSGRESNNDTGLTITAKNVEVNDKITSKLT 1140
Qy 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTSKSGKIEA 1200
Db 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTSKSGKIEA 1200
Qy 1201 KSGEANVTSAITGTTISGNTVSVNATAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
Db 1201 KSGEANVTSAITGTTISGNTVSVNATAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
Qy 1261 GSSITSTKGOVDLLAONGSIAGSINAANVTLLNTTGLTTTVAGSDIKATSGTLVINAKDAK 1320
Db 1261 GSSITSTKGOVDLLAONGSIAGSINAANVTLLNTTGLTTTVAGSDIKATSGTLVINAKDAK 1320
Qy 1321 LNGDASGDSSTEVNAVNASGSGVTAATSSSVNITGDLNTVNGLNISKDGRNTVRLRGKE 1380
Db 1321 LNGDASGDSSTEVNAVNASGSGVTAATSSSVNITGDLNTVNGLNISKDGRNTVRLRGKE 1380
Qy 1381 IEVKYIQPGVASVEVIEAKRVLEKVKDLSDEERETLAKGVSAVRPEPNNTITVANTON 1440
Db 1381 IEVKYIQPGVASVEVIEAKRVLEKVKDLSDEERETLAKGVSAVRPEPNNTITVANTON 1440
Qy 1441 EFTTRPSSQVIIESEKACFPSSGNGARVCTNVADGQP 1477
Db 1441 EFTTRPSSQVIIESEKACFPSSGNGARVCTNVADGQP 1477

RESULT 6
AAW30292
ID AAW30292 standard; Protein; 1601 AA.
XX
AC AAW30292;
XX
DT 14-APR-1998 (first entry)
XX
DE Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX

```

```

KW Non-typeable Haemophilus; high molecular weight surface protein;
XX HMW4; immunogen; vaccine; otitis media.
OS Haemophilus influenzae strain 5.
FH Key Location/Qualifiers
FT Misc-difference 372 /note= "encoded by TCT"
FT Misc-difference 400 /note= "encoded by AAT"
FT
XX WO9736914-A1.
XX
XX 09-OCT-1997.
XX
XX 01-APR-1997; 97WO-US04707.
XX
XX 01-APR-1996; 96US-0617697.
XX
XX (BARE/) BARENKAMP S J.
XX
XX Barenkamp SJ;
XX
XX WPI; 1997-503038/46.
XX
XX N-PSDB; AAT90993.
XX
XX High molecular weight proteins of non-typeable Haemophilus
XX influenzae - useful for vaccine production
XX
XX Claim 1: Page 97-102; 183pp; English.
XX
XX This protein comprises the high molecular weight surface protein
XX HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
XX has the immunological ability to protect against disease caused by
XX a non-typeable Haemophilus strain and is characterised by at least
XX one surface-exposed B-cell epitope that is recognised by monoclonal
XX antibody A06. The HMW4 amino acid sequence was deduced from an
XX isolated hmw4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
XX AAW30294) and HMW3 (see AAW30291) have also been identified. A
XX conjugate comprising HMW4 linked to an antigen haptan or
XX polysaccharide, and a synthetic peptide of 6-150 amino acids
XX corresponding to at least protective epitope of HMW4 are also
XX claimed. HMW proteins, conjugates and peptides can be used in
XX vaccines, as immunogens for preparation of antibodies and as
XX antigens for detection of these antibodies.
XX
XX Sequence 1601 AA:
XX
XX Query Match 71.3%; Score 5284; DB 18; Length 1601;
XX Best Local Similarity 68.4%; Pred. No. 3.4e-243;
XX Matches 1100; Conservative 136; Mismatches 231; Indels 142; Gaps 16;
Qy 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPKVRHLAKPLSALLSLGVT 60
Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPKVRHLAKPLSALLSLGMA 60
Qy 61 SIPOSVLASGLQGVVHGVTATMVDGNTIIRNSVDAILNKKOFNIDQNMVQFLQENN 120
Db 61 SIPOSVLASGLQGVVHGVTATMVDGNTIIRNSVDAILNKKOFNIDQNMVQFLQESS 120
Qy 121 NSAVFNRTVSTQISQKGLDLSNGOVFLINPNTGKDAINTNGTASTLDSINENIK 180
Db 121 NSAVFNRTVSTQISQKGLDLSNGOVFLINPNTGKDAINTNGTASTLDSINENIK 180
Qy 181 ARNFTFQTKDALKAEIVNGLITVGVKDSVNLIGGKVKNEGVSIVNGSGSISLLAGKIT 240
Db 181 ARNFTFQTKDALKAEIVNGLITVGVKDSVNLIGGKVKNEGVSIVNGSGSISLLAGKIT 240
Qy 241 ISDIINPTITYSIAAPENEAIVNLGDIKAGGNINVRATIRNOCKLSADSVSKDGNIV 300
Db 241 ISDIINPTITYSIAAPENEAIVNLGDIKAGGNINVRATIRNOCKLSADSVSKDGNIV 300
Qy 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKGTAVIDLSGEGGETYLGGBERGE 360

```



Db	1372	TLTAE\$KGLTQTQGS\$IT\$SNGQTLLTAKD\$SIAGNNAANVTLLNTTCTLLTTTGDSKINA	1431
QY	1308	TSGLTVINAKDAKLN\$GDS\$PEYNAVNASG\$SVT\$AT\$SSVNIITGDLNTVNGLNIIS	1367
Db	1432	TSGLTITINAKDAKLDCA\$GDRTVVNATNASG\$GWNVTAKT\$SSVNIITGDLNTVNGLNIIS	1491
QY	1368	KDGRNTVRLRGKEIEVKYIQPGVASVEEIEAKRVLEKVKDLSDEERETTLAKLGV\$AVRF	1427
Db	1492	ENGRNTVRLRGKEIDVKYIQPGVASVEEIEAKRVLEKVKDLSDEERETTLAKLGV\$AVRF	1551
QY	1428	VEPNNTITVNTQNEFTTRESSOVIITSE\$KACF\$SSNGNGARVCTNVAD\$GQ	1476
Db	1552	VEPNNAITVNTQNEFTT\$PSQVITSE\$KACF\$SSNGNGARVCTNVAD\$GQ	1600

AAB01849
ID AAB
XX AC AAB
XX XX
DT 11-
XX DE Hae
XX KW Mat
KW KW non
KW KW rec
KW KW huma
KW KW det

OS Haer

XX  
PN WO20[illegible]

PD 13-A  
XX

07-C

XX PR 07-C

PR 08-D

XX  
PA (CONXX  
DTXX  
FI  
SOOT

DR	WPI;
DR	N = 100

XX  
N-PS

PT	Nucl
PT	Nucl

protd  
protd

XX  
XX

XX  
PS  
CJAL

CC The inf

CC	expr.
CC	int

CC CC prom.

CC to a  
CC infl

CC	clus
CC	clus

nmwB and nmwB

CC  
resp

the prote

CC  
matur

and strain

vector  
prot

cell-  
proce

1

CC humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HmW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, CC HmW proteins and/or HmW peptides. The nucleotide sequences encoding the CC HmW proteins can be used to isolate and clone hmw genes from other CC non-typeable strains of Haemophilus via hybridisation reactions. The CC -presentable strains of Haemophilus via hybridisation reactions. The CC -presentable strains of Haemophilus via hybridisation reactions. The CC strain of H. influenzae.

xx Sequence 1036 AA;

Query Match 69.7%; Score 5164; DB 21; Length 1036;  
Best Local Similarity 99.5%; Pred. No. 1e-237;  
Matches 1030; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 443 DNVSNIAEDPLENNTGINDPEFTGTGEADPKKNSLKTTLTNTTISNYLKNWMTNITA 502  
Db 2 DDVTIEADPLRNTGINDPEFTGTGEADPKKNSLKTTLTNTTISNYLKNWMTNITA 61  
Qy 503 SRKLTVNSSINIGNSHLILHSGGQGGVQIDGDIITSGKGLTIYSGGWDVHKNTILD 562  
Db 62 SRKLTVNSSINIGNSHLILHSGGQGGVQIDGDIITSGKGLTIYSGGWDVHKNTILD 121  
Qy 563 OGFLNITAAVAPEGGNKKARDANAKIVAQGVTTITGEGKDFRANNSVLTGKGLNII 622  
Db 122 OGFLNITAAVAPEGGNKKARDANAKIVAQGVTTITGEGKDFRANNSVLTGKGLNII 191  
Qy 623 SSVNLTNLSGTTINISGNTITINQTRKNTSYQWTSKSHWNSALMFTGANFTIKYI 682  
Db 182 SSVNLTNLSGTTINISGNTITINQTRKNTSYQWTSKSHWNSALMFTGANFTIKYI 241  
Qy 683 SSNSKGLTQYRSSAGVNFNGVNGMSFNLKEGAKVNFKLKPNENMNTSKPLPIRLANI 742  
Db 242 SSNSKGLTQYRSSAGVNFNGVNGMSFNLKEGAKVNFKLKPNENMNTSKPLPIRLANI 301  
Qy 743 TATGGGVFFDIYAHNSGRGAELKMSINISNGANFTLNSHVRGDDAFKINKDLTINATN 802  
Db 302 TATGGGVFFDIYAHNSGRGAELKMSINISNGANFTLNSHVRGDDAFKINKDLTINATN 361  
Qy 803 SNFSLRQTKDQFYDGYARNAINSTYINISILGGNVTLGGQNSSSSITGNITIEKAANVTLE 862  
Db 362 SNFSLRQTKDQFYDGYARNAINSTYINISILGGNVTLGGQNSSSSITGNITIEKAANVTLE 421  
Qy 863 ANNAPQONIRDRYIKGLSLVNGSLTGTENADIKGNLTISESATPKGTRDTLTNITGN 922  
Db 422 ANNAPQONIRDRYIKGLSLVNGSLTGTENADIKGNLTISESATPKGTRDTLTNITGN 481  
Qy 923 FTNNGTAENITQGVVKGNGVNDGDLNITTHAKRNORSIIGGDIINKKSLNITDSND 982  
Db 482 FTNNGTAENITQGVVKGNGVNDGDLNITTHAKRNORSIIGGDIINKKSLNITDSND 541  
Qy 983 AEIQIGNISOKENLTISDKINITYITIKKIDGESSDTSANLTITKTELKLT 1042  
Db 542 AEIQIGNISOKENLTISDKINITYITIKKIDGESSDTSANLTITKTELKLT 601  
Qy 1043 EDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKVTTFNNVKDSKISADGHNVTLNSK 1102  
Db 602 EDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKVTTFNNVKDSKISADGHNVTLNSK 661  
Qy 1103 VRTSSNGGRESNDNDTGLTITARNVENVKDIITSLKTVNITASEKVTITAGSTINATNG 1162  
Db 662 VRTSSNGGRESNDNDTGLTITARNVENVKDIITSLKTVNITASEKVTITAGSTINATNG 721  
Qy 1163 KASITTKTGDLSGTSIGNTVSVATVLTITKSGSKIEAKSGEANTVATGTTIGTISGNT 1222  
Db 722 KASITTKTGDLSGTSIGNTVSVATVLTITKSGSKIEAKSGEANTVATGTTIGTISGNT 781  
Qy 1223 VNVATANAGDLVNGAEINATEGAATLTATGNTLTTEAGSSITSTKQVDVLLAONGSIAG 1282  
Db 782 VNVATANAGDLVNGAEINATEGAATLTATGNTLTTEAGSSITSTKQVDVLLAONGSIAG 841  
Qy 1283 SINAAVNLTNTTGLTTVAGSDIKATSGTLVINAKDAKNGDASGDSTEVAVNASGSGS 1342  
Db 1342 SINAAVNLTNTTGLTTVAGSDIKATSGTLVINAKDAKNGDASGDSTEVAVNASGSGS 1021

Db 842 SINAAVNLTNTTGLTTVAGSDIKATSGTLVINAKDAKNGDASGDSTEVAVNASGSGS 901  
Qy 1343 VTAATSSSVNTGDLNTVNGLNIIISKDGRNVRRLRGKEIEVKYIOPGVASVEEVEAKRV 1402  
Db 902 VTAATSSSVNTGDLNTVNGLNIIISKDGRNVRRLRGKEIEVKYIOPGVASVEEVEAKRV 961  
Qy 1403 LEKVKDLSEDERETLAKLGVSARVRENNITVTQNEFTTRPSSOVIISGKACFSSG 1462  
Db 962 LEKVKDLSEDERETLAKLGVSARVRENNITVTQNEFTTRPSSOVIISGKACFSSG 1021  
Qy 1463 NGARVCTNVADDGQP 1477  
Db 1022 NGARVCTNVADDGQP 1036

# RESULT 8

AAAR41723

ID AAR41723 standard; Protein; 1536 AA.

XX AC AAR41723;

XX DT 26-APR-1994 (first entry)

XX DE High molecular weight protein 1 (HmW1).

XX KW HmW; high molecular weight protein; virus; vaccine; influenza;

XX KW epitope; immunity; haemophilus influenzae.

XX OS Haemophilus Influenzae.

XX PN WO9319090-A.

XX PD 30-SEP-1993.

XX PF 16-MAR-1993; 93WO-0502166.

XX PR 16-MAR-1992; 92GB-0005704.

XX PA (BARE/) BARENKAMP S J.

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Barenkamp SJ;

XX DR WPI; 1993-320683/40.

XX DR N-PSDB; AAQ49506.

XX PT High molecular weight surface proteins - of non-typeable

XX PT haemophilus which exhibit immunogenic properties

XX PS Claim 3; Figure 2; 100pp; English.

CC The isolation and purification of the high molecular weight protein enables the identification of the major protective epitopes of the protein by conventional epitope mapping. These epitopes can then be synthesised using standard techniques and incorporated into fully synthetic or recombinant vaccines.

XX SQ Sequence 1536 AA;

Query Match 67.1%; Score 4967; DB 14; Length 1536;  
Best Local Similarity 67.3%; Pred. No. 3.9e-228;  
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKESKPARMKVRHLAKPLSAMLISLGV 60

Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKESKPARMKVRHLAKPLSAMLISLGV 60

Qy 61 SIFQSVLASGLOGMDVVGHTATMOVGDKTIIRNSDAIINWKFQNFIDQNMVQFLOENN 120

Db 61 SIFQSVLASGLOGMDVVGHTATMOVGDKTIIRNSDAIINWKFQNFIDQNMVQFLOENN 120

Qy 121 NSAVFNRTVSNQISQLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180

Db 121 NSAVFNRTVSNQISQLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180

Db 121 NSAVENRVTSNQISQLKGLDSNGQVFLNPNNGIITIGKDAIINTNGFTASTLIDISNENIK 180  
 QY 181 ARNFTFEQTKKALAEIVNHGLITVKGDSVNLIGKVKNEGIVSVNGSGSISLLAQKIT 240  
 Db 181 ARNFTFEQTKKALAEIVNHGLITVKGDSVNLIGKVKNEGIVSVNGSGSISLLAQKIT 240  
 QY 241 ISDIINPTITYSIAAPENAEVNLGDFPAKGGNINVRATIRNOGKLSADSVSKDKNIV 300  
 Db 241 ISDIINPTITYSIAAPENAEVNLGDFPAKGGNINVRATIRNOGKLSADSVSKDKNIV 300  
 QY 301 LSAKEGEAEIGVISAQNOQAGKGLMTITGDKVTLTKTCAVIDLDSKGEGETYLGDERGE 360  
 Db 301 LSAKEGEAEIGVISAQNOQAGKGLMTITGDKVTLTKTCAVIDLDSKGEGETYLGDERGE 360  
 QY 361 KGNGIOLAKTSLKSGSTINVSKEGKGFAIWMGDIALIDGNINAQSGDIAKTGGFVET 420  
 Db 361 KGNGIOLAKTSLKSGSTINVSKEGKGFAIWMGDIALIDGNINAQSGDIAKTGGFVET 420  
 QY 421 SGHDLFIKDNAIVDAKEWLLDFDWNVINAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479  
 Db 421 SGHDLFIKDNAIVDAKEWLLDFDWNVINAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479  
 QY 480 KTTLTNTTISNLYKNAMTMTTASRKLTVNSINIGSNLSHLHSKGGGVOIDGDIT 539  
 Db 479 KTTLTNTTLESILKGFVNITANQRIYVNSINL-SNGSLTLWSEGRSGGVLEINNDIT 537  
 QY 540 ----SKGGLTIYSGGWVHVHKNITLD-QGFLNITA-ASVAFEGGNKARDAANAIVAQ 593  
 Db 538 TGDPTRGANLTIIYSGGWVHVHKNISLQAGNINITAKDIAFEKGSQV-----ITGQ 590  
 QY 594 GTVITIGEGKDFRANNVSLNCTGKGLNIISVNN--LTHNLSTINTSGNITINQTRK 650  
 Db 591 GTIIT-SGNQKGRFNNVSLNCTGKGLNIISVNN--LTHNLSTINTSGNITINQTRK 649  
 QY 651 NTS-YWOTSHDHNWVSALETCANFTF-IKYISSNKGTLTYRSGAGVNFNGVAGNM 708  
 Db 650 NESGYDFKGRTYNLTSLNVSSEGEENLTIDSRGSDSAGTLTQYNLNGISF--NKDT 706  
 QY 709 SFNLKEGAKVNFKLKPNENMTSKPLP-IRPLANITATGGSGVFFDIYANHSQ---RGAE 764  
 Db 707 TFNVERNARVNDIKAPIGINKYSLNVSFNGNISVSGGSDFTLLASSNSVQTPCV 766  
 QY 765 LKMEINISGANFTLNSHVGRDPAFKINKDLTINATNSNFSLRQTKDDPFDGYARNAIN 824  
 Db 767 INSKYFNVSTGSSLRFTSGTKTGFSTERKDLTLNATGGNITLLQVEGT--DGMIGKIV 824  
 QY 825 STYNISLGGNVTLGGQNSSTTGNITIEKAAVNTLEANNAPQOIRDRVILKGLSLV 884  
 Db 825 AKKNITFEGGNITFGSRKAVTEIEGNTINNANVTILIGSDFDNHQ--KPLTIKKDVIIN 882  
 QY 885 NGSLSLGTENADIKGNTLISFATFKGTRDTLNTGNTNNGTAETINITQGVVKGNGVT 944  
 Db 883 SGNLTAGNVINAGNLTVESNFAKAITNFTFNVGGLFDKNGNSNISIAKGGARFKDID 942  
 QY 945 NQDLNITTHAKNRQSIIGGDIINKGSLNITDSNDAETIQIGNTSQKEGNTLISDDK 1004  
 Db 943 NSKLSITSSSYTYRIISGNITNKNKDLNINNEGSDTEWQIGDVSQKEGNTLISDDK 1002  
 QY 1005 INTKQITIKKGDGEDSSDATSNANLTIKTRELKLTEDLSISGPNKAEITAKDGRDLT 1064  
 Db 1003 INITKQITIKAGVDGNSDATNANLTIKTRELKLTQDLNISGPNKAEITAKDGSDLT 1062  
 QY 1065 IGSNDGNSCAEAKTYFNNVKDSKISADGHNTVLSKVYTSNGGREGNSNDTGLTI 1124  
 Db 1063 IGTNSAD-GTNNAKVTFNQVKDSKISADGHKVTLSKVYTSNGGREGNSNDTGLTI 1121  
 QY 1125 TAKNVEYKDTISLTKVNITA-SEKVTITTAGSTINAFNGKASITTKT----- 1170  
 Db 1122 DAKNVTNNITSHKAVISATSEITTKGTINATGTVNVEITAGTGSILGIESSGS 1181  
 QY 1171 -----GDISGETISGNT 1181  
 Db 1182 VTLTATEGALAVSNISGNTVTVTANSALTTLAGSTIKGTESVTTSSQSGDITGTSGET 1241

QY 1182 VSVSATVDLTTRKSGKIEAKSGEANTVSATGTIGGTISGNTVNTVNTAGDLTVNGAEIN 1241  
 Db 1242 VEVKATESLTTOSNSKIKATTEANTVSATGTIGGTISGNTVNTVNTAGDLTVNGAEIN 1301  
 QY 1242 ATEGAATLTATGNTLTTEAGSSITSTKGQVLLLAQNGSIAGSINAANVTLNTTGLTTVA 1301  
 Db 1302 ATEGAATLTSSGKLTTEASSHITSAKQVNLQAQDSGVAGSINAANVTLNTTGLTTVK 1361  
 QY 1302 GSDIKATSGTLVINAKDAKLGDSGDSSTEVNAVNASGSGSVTAATSSVNTTGDINTVN 1361  
 Db 1362 GSNINATSGTLVINAKDAELGALGNHTVVNATNANGSGSVIATTSRVNTTGDITIN 1421  
 QY 1362 GLNIISKGRNTRVRURGEIEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLG 1421  
 Db 1422 GLNIISKNGINTVLKGVKIDVKYIQPGIASVDEIEAKRVLEKVKDLSDEERETLAKLG 1481  
 QY 1422 VSAVREVPNNITVTNTQNEFTTREPSSQVIISSEKACFSSNGARVCTNVADDDQ 1476  
 Db 1482 VSAVREVPNNITVTNTQNEFATPLSRIVISEGRACFSNSDGTATVCVNIADNGR 1536

## RESULT 9

AA41732  
 ID AAR41732 standard; Protein; 1529 AA.

XX AAR41732;

XX 26-APR-1994 (first entry)

XX High molecular weight protein 4 (HMM4).

XX HMW; high molecular weight protein; virus; vaccine; influenza;

XX epitope; immunity; haemophilus influenzae.

XX Haemophilus influenzae.

XX WO9319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993; 93WO-US02166.

XX 16-MAR-1992; 92GB-0005704.

XX (BARE/) BARENKAMP S J.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

XX N-PSDB; AAQ49511.

XX High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties

PS Claim 6; Figure 10; 100pp; English.

XX The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.

XX Sequence 1529 AA;

Query Match 66.9%; Score 4957; DB 14; Length 1529;

Best Local Similarity 67.2%; Pred. No. 1.2e-227;

Matches 103; Conservative 135; Mismatches 227; Indels 142; Gaps 16;

QY 73 GMDVHGHTATMQVDGKNTIIRNSVDAIINNKQFNIDQNEVQFLOENNSAVFNRTSQ 132

Db 1 GMSVHGHTATMQVDGKNTIIRNSVDAIINNKQFNIDQNEVQFLOENNSAVFNRTSDQ 60

QY 133 ISQLKGLDSNGOVFLINPGITIGKDAIINTNGFTASTLIDISNENIKARNFTFEOTKDK 192  
 DB 133 ISQLKGLDSNGOVFLINPGITIGKDAIINTNGFTASTLIDISNENIKARNFTFEOTKDK 120  
 QY 193 ALAEIVNHLITVKGOSVNLIGKVKNEGVISVNGGSSLLAGOKITISDIINPIITYS 252  
 DB 121 ALAEIVNHLITVKGOSVNLIGKVKNEGVISVNGGSSLLAGOKITISDIINPIITYS 180  
 QY 253 IAAPENAVLGDIPAKGGNINVRATIRNOGKLSADSVKSKSGNIVLSAKGEAEIGG 312  
 DB 181 IAAPENAVLGDIPAKGGNINVRATIRNOGKLSADSVKSKSGNIVLSAKGEAEIGG 240  
 QY 313 VISAQNOQAKGGKIMITGDKVTLTKTGAVIDLSGKEGETYLGDERGEGKNGIOLAKKTS 372  
 DB 241 VISAQNOQAKGGKIMITGDKVTLTKTGAVIDLSGKEGETYLGDERGEGKNGIOLAKKTS 300  
 QY 373 LEKSTINVSKEKGGFAIVMGDIALIDGNINAQGGSDIAKTGGFVETSCHDLFIKDNAI 432  
 DB 301 LEKSTINVSKEKGGFAIVMGDIALIDGNINAQGGSDIAKTGGFVETSCHDLFIKDNAI 359  
 QY 433 VDAREWLLDPNVINAEDPLENNTGIDNFTGTGEASDPKKNSELKTLTNTTISNYL 492  
 DB 360 VDAREWLLDPNVINAEDPLENNTGIDNFTGTGEASDPKKNSELKTLTNTTISNYL 419  
 QY 493 KNAWMTNITASRKLTVNSNIGNSHILHLSKQGGVQIDGDTIS-KGGNLTIVSGG 551  
 DB 420 RRGSVYNTANRIYVNSSINL-SNGSLTLTKRD---GVKINGDITSNENGLTIKAGS 475  
 QY 552 WDVHKNITDQGLNITAA-SVAFEGGNKARDAANAKIVAOGTVITITGEGKDFRANV 610  
 DB 476 WDVHKNITDQGLNITAA-SVAFEGGNKARDAANAKIVAOGTVITITGEGKDFRANV 535  
 QY 611 SLNGTGKGLNITISSVNNLTHLSGTINISGNITITQTRKNTSYWOTSHDSHNVNLSALNL 670  
 DB 536 SINGTGKGLNITISSVNNLTHLSGTINISGNITITQTRKNTSYWOTSHDSHNVNLSALNL 595  
 QY 671 ETGANFTPIKYI--SSSKSGLTTOYRSSAGVNFNGVNGNMSFNLEKAGVNFKLKPNENM 728  
 DB 596 NTVOKFTPIKVDGSGNSQDLRSRRSFAGVHENGIGGKTNFNIGANAKALFKLPNAAT 655  
 QY 729 NTSKPLPRLFLANTATG--GSVFEDYIYANHSRGRGELKMSSEINISGANFTLNHVRG 786  
 DB 656 DPKKELPITFNANITATGSDSSVMEDIHANLTSRAAGINMDSINITGGLDFSIISHNRN 715  
 QY 787 DDAFKINKDLTINATNSNFSLOTKDDFYDGYARNAINSTYNIISILGNYTLGGONSSS 846  
 DB 716 SNAFEIKDLTINATNSNFSLOTKDDFYDGYARNAINSTYNIISILGNYTLGGONSSS 775  
 QY 847 ITGNITIEKAANVLEANNAPNOQNIORRVIKLGLSLLVNGSLTGENADIKNLTISES 906  
 DB 776 ITGNITIEKAANVLEANNAPNOQNIORRVIKLGLSLLVNGSLTGENADIKNLTISES 835  
 QY 907 ATFKGKTRDTLNTGNFTNGTAEINITOGVVKL-GNVYNDGDLNITHAKRNQRSIIGG 965  
 DB 836 STFKGEASDNLTITFTFNGTANINIKQGVYKLGQDIINKGGLNITNASGTOKTIING 895  
 QY 966 DIINKKGLNITDSNNDAEIQIGGNISQKGNLTISSDIKNITKQITIKKIDGEDSSD 1025  
 DB 896 NITNEKGLNITKIKADAEIQIGGNISQKGNLTISSDIKNITKQITIKKIDGEDSSD 955  
 QY 1026 ATSNANLTITKTELKTLDELSTGSPNKAPITAKDGRDITGNSNDGSGAEAKTVTFNNV 1085  
 DB 956 EAENANLTITKTELKTLDELSTGSPNKAPITAKDGRDITGNSNDGSGAEAKTVTFNNV 1013  
 QY 1086 KOSKISADGNVNTLSKVYKTSNGSGRESNDSNDTGLTITAKNVEYKNDITSLAKTVAI-- 1143  
 DB 1014 KOSKISADGNVNTLSKVYKTSNGSGRESNDSNDTGLTITAKNVEYKNDITSLAKTVAI-- 1071  
 QY 1144 -----TASEKVTTTAGSTINATNGK 1163  
 DB 1072 AAGNVTTKEGTTINATGTSVEVTAQNGTIKGNITTSQNVTVTATENLVTNAVINATSGT 1131

QY 1164 ASITTKTGDISG-----TISGNTSVSVS-----ATVD---LTTKSGSKIEAK 1201  
 DB 1132 VNISTKTGDIKGIESTSGNVNITASGNTLKVSNITGQDVTVTADAGALTTTAGSTISAT 1191  
 QY 1202 SGEANVTSGT-----TIG--- 1215  
 DB 1192 TGNANITTKTGIDINGKVBESSGSVTLVATGATLVAGNISGNVTVTITADSGKLTSTVGSTI 1251  
 QY 1216 -----CTISGNTVNTANAGDLTVNGCAEINATGAATLTATGNTLTTE 1259  
 DB 1252 NGTNSVTTSSQSDTEGISTISGNTVNTASTGTDTIGNSAKVEAKNGAATLTAESGKLTQ 1311  
 QY 1260 AGSSTSTKGVDDLLAONGSIAGSINAANVTLLTNTTGTITTVAGSDIKATSGTLVINAKDA 1319  
 DB 1312 TGSSITSSNGQTLTAKDSSIAAGNINAANVTLLTNTTGTITTVAGSDIKATSGTLVINAKDA 1371  
 QY 1320 KLNGDASGSTEVAVNAVNASGSGSVTAATSSSVNITGDLTVNGLNIISKDGRNTVRLRGK 1379  
 DB 1372 KLDGAASGDRTVVNTATNAGSGNVTAKTSSSVNITGDLTVNGLNIISKDGRNTVRLRGK 1431  
 QY 1380 EIEVKYIQGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFEVPEPNTITVNTQ 1439  
 DB 1432 EIDVKYIQGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFEVPEPNTITVNTQ 1491  
 QY 1440 NEFTTRPSSQVITISEGKACFSSGNGARVCTNVADDGQ 1476  
 DB 1492 NEFTTRPSSQVITISEGKACFSSGNGARVCTNVADDGQ 1528

## RESULT 10

AAW30293  
 ID AAW30293 standard; Protein; 1536 AA.  
 XX  
 AC AAW30293;  
 XX  
 DT 14-APR-1998 (first entry)  
 XX  
 DE Non-typeable Haemophilus high mol.wt. surface protein HMW1.  
 KW Non-typeable Haemophilus; high molecular weight surface protein;  
 KW HMW1; hmwlA gene; immunogen; vaccine; otitis media.  
 OS Haemophilus influenzae strain 12.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 4 /note= "encoded by CTA"  
 FT Misc-difference 98 /note= "encoded by GAT"  
 FT Misc-difference 363 /note= "encoded by AAG"  
 XX  
 PN W09736914-AL.  
 XX  
 PD 09-OCT-1997.  
 XX  
 PF 01-APR-1997; 97WO-0504707.  
 XX  
 PR 01-APR-1996; 96US-0617697.  
 XX  
 PA (BARE/) BARENKAMP S J.  
 XX  
 PI Barenkamp SJ;  
 XX  
 DR WPI: 1997-503038/46.  
 XX  
 DR N-PSDB; AAT90994 and AAT90996.  
 XX  
 PT High molecular weight proteins of non-typeable Haemophilus  
 PT influenzae - useful for vaccine production  
 XX  
 PS Claim 7; Page 66-70; 183pp; English.  
 CC This protein comprises the high molecular weight surface protein

CC *HMW1* (125 kDa) of non-typeable *Haemophilus influenzae* strain 12 that  
 CC has the immunological ability to protect against disease caused by a  
 CC non-typeable *Haemophilus* strain and is characterised by at least  
 CC one surface-exposed B-cell epitope that is recognised by monoclonal  
 CC antibody AD6. The *HMW1* amino acid sequence was deduced from the  
 CC *HMW1* gene sequence (see AA790994 and AA790996). The expressed protein  
 CC is truncated, starting at residue 442 of the full-length gene  
 CC product. *HMW2* (see AA302094), *HMW3* (see AA302091) and *HMW4* (see  
 CC AA302092) have also been identified. A conjugate comprising *HMW1*  
 CC linked to an antigen, hapten or polysaccharide, and a synthetic  
 CC peptide of 6-150 amino acids corresponding to at least protective  
 CC epitope of *HMW1* are also claimed. *HMW* proteins, conjugates and  
 CC peptides can be used in vaccines, as immunogens for preparation of  
 CC antibodies and as antigens for detection of these antibodies.  
 CC  
 SQ Sequence 1536 AA;

Sequence	1536 AA:
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120
121	121
122	122
123	123
124	124
125	125
126	126
127	127
128	128
129	129
130	130
131	131
132	132
133	133
134	134
135	135
136	136
137	137
138	138
139	139
140	140
141	141
142	142
143	143
144	144
145	145
146	146
147	147
148	148
149	149
150	150
151	151
152	152
153	153
154	154
155	155
156	156
157	157
158	158
159	159
160	160
161	161
162	162
163	163
164	164
165	165
166	166
167	167
168	168
169	169
170	170
171	171

Query Match	Score	DB	Length
66.9%	4956	18	1536

Matches	1046;	Conservative	146;	Mismatches	227;	Indels	98;	Gaps
QY	1	MNKYRLKFKSR	NALVAVSELARGCDHST	KEGSEKPARMKVRHLAL	KPLSAML	LSL	SLGVT	60
Db	1	MNKYRLKFKSR	NALVAVSELARGCDHST	KEGSEKPARMKVRHLAL	KPLSAML	LSL	SLGVT	60
QY	61	SIQSVLASL	QGMVDVHGVTATMOVDG	NKTIIRNSVDAI	INNKQFNI	DONEMVQ	FOLENN	120
Db	61	SIQSVLASL	QGMVDVHGVTATMOVDG	NKTIIRNSVDAI	INNKQFNI	DONEMVQ	FOLENN	120
QY	121	NSAVFNRVTS	NIQSILKGLD	NSGVQVFLINP	NGITIGKDAI	INTNGFTAST	LIDS	180
Db	121	NSAVFNRVTS	NIQSILKGLD	NSGVQVFLINP	NGITIGKDAI	INTNGFTAST	LIDS	180
QY	181	ARNFTFEQ	KDKALAEIVNHGL	ITVKGDSVNLIG	GVKN	EGVISVNG	GSIS	240
Db	181	ARNFTFEQ	KDKALAEIVNHGL	ITVKGDSVNLIG	GVKN	EGVISVNG	GSIS	240
QY	241	ISDIINPT	ITYSTIAAP	ENEAVNLG	DFAGGNNINRAA	TIRNOGKL	SADSVSK	300
Db	241	ISDIINPT	ITYSTIAAP	ENEAVNLG	DFAGGNNINRAA	TIRNOGKL	SADSVSK	300
QY	301	LSAKEGEAE	IGGVISAQNOOAK	GKGLMITGDKV	TLTKGAV	IDL	SGKEG	420
Db	301	LSAKEGEAE	IGGVISAQNOOAK	GKGLMITGDKV	TLTKGAV	IDL	SGKEG	420
QY	361	GNKGIOLAK	TSLEKGSTIN	VSKEKGGPAI	VWGDIAL	LDG	IN	480
Db	361	GNKGIOLAK	TSLEKGSTIN	VSKEKGGPAI	VWGDIAL	LDG	IN	480
QY	421	SGHDLFI	IKDNAIVDAK	EWLLDPDNV	SINAEDPLFN	NTG	IN	540
Db	421	SGHDLFI	IKDNAIVDAK	EWLLDPDNV	SINAEDPLFN	NTG	IN	540
QY	480	KTTLNTNT	ISLYLKN	AWTNIT	ASKRLTV	NS	INTG	600
Db	479	KTTLNTNT	LESILK	GGTFV	YNITANQ	RIYV	NS	600
QY	540	---	SKGGNLT	ITYSGWVDV	VHKNITLD	-QGFL	NITA	660
Db	538	TGDDTR	GANLT	ITYSGWVDV	VHKNITLD	-QGFL	NITA	660
QY	594	GPVTIT	GEKGF	FRANV	SLNGT	GKGLN	IT	720
Db	591	GTIT	-SGN	KGF	FRANV	SLNGT	GKGLN	720
QY	651	NTS	-YWQ	TSHD	SHNV	SALN	LET	780
Db	650	NESGYD	KFG	TYW	NLT	SLN	YSE	780
QY	709	SFNL	KEG	AV	FKL	P	KN	840
Db	707	TEN	VER	NAR	V	AF	D	840

RESULT 11  
AAR63505

ID AAR63505 standard; Protein; 1536 AA.

XX AAR63505;

XX  
DT 25-JUN-1995 (first entry)

DE Haemophilus high molecular weight protein HMWL. XX

XX	
KW	High molecular weight protein; HMW1; protective vaccine; otitis;
KW	sinusitis; bronchitis; Hib.

XX  
OS  
Haemophilus.XX  
PN W09A21290-A

XX

10 43 SEP 1994.  
XX  
PF 15-MAR-1994; 94WO-US02550.

XX PR 16-MAR-1993; 93US-0038682.  
 XX PA (BARE/) BARENKAMP S J.  
 XX PA (SGEM/) ST GEME J W.  
 XX X  
 XX PI Barenkamp SJ, St GEME JW;  
 XX X  
 XX DR WPI; 1994-316665/39.  
 XX DR Q-PSDB; Q72293.  
 XX X  
 XX PT New immunogenic high mol. wt. proteins of non typeable  
 XX PT Haemophilus - useful in protective vaccines  
 XX X  
 XX PS Claim 2; Page 31; 127pp; English.  
 XX X  
 CC CC The HmW1 protein encoded by this sequence is useful in a vaccine to  
 CC CC protect against disease caused by non-typeable Haemophilus which are  
 CC CC not controlled by H. influenzae type b (Hib) vaccines. The encoded  
 CC CC protein can also be used as a carrier for protective Hib  
 CC CC polysaccharide (in a conjugate vaccine against meningitis) or for  
 CC CC other antigens, haptens, etc.  
 XX X  
 SQ Sequence 1536 AA;

Query Match 66.7%; Score 4943; DB 15; Length 1536;  
 Best Local Similarity 67.1%; Pred. No. 5.4e-227;  
 Matches 1043; Conservative 147; Mismatches 267; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKGESEKPARMKVRHLALKPLSALLSLGVT 60  
 Db 1 MNKIYRLKFKSRNALVAVSELARGCDHSTKGESEKPARMKVRHLALKPLSALLSLGVT 60  
 Qy 61 SIPOSVLASGQGMVHVGTATMOVGDKNTIIRNSVDIIINWKOFNDQEMVQFLOENN 120  
 Db 61 SIPOSVLASGQGMVHVGTATMOVGDKNTIIRNSVDIIINWKOFNDQEMVQFLOENN 120  
 Qy 121 NSAVFNRTVSNQISQLKGLDSNGOVFLINPGITIGCKDAIINTNGFTASTLDISNENIK 180  
 Db 121 NSAVFNRTVSNQISQLKGLDSNGOVFLINPGITIGCKDAIINTNGFTASTLDISNENIK 180  
 Qy 181 ARNTEFQTKDKALAEIVNHLGIIIVGKDSVNLGKGVKNEGVLSVNGSGISLLAGOKIT 240  
 Db 181 ARNTEFQTKDKALAEIVNHLGIIIVGKDSVNLGKGVKNEGVLSVNGSGISLLAGOKIT 240  
 Qy 241 ISDIINPTIYISIAAPENEAIVLGDIFAKGGINIVRAATIRNOQKLSADSVSKDKSNIV 300  
 Db 241 ISDIINPTIYISIAAPENEAIVLGDIFAKGGINIVRAATIRNOQKLSADSVSKDKSNIV 300  
 Qy 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGETYLGGERGE 360  
 Db 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGETYLGGERGE 360  
 Qy 361 GKNGIQLAKKTSLEKSGTINVSQKGEKGFIAVWGDIALIDGNINAQSGDIATKGGFVET 420  
 Db 361 GKNGIQLAKKTSLEKSGTINVSQKGEKGFIAVWGDIALIDGNINAQSGDIATKGGFVET 420  
 Qy 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNVNAEDPLFNNTGGINDEFTPTG-BASDPKKNSEL 479  
 Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNVNAEDPLFNNTGGINDEFTPTG-BASDPKKNSEL 479  
 Qy 480 KTLTNTTISNLYKNWMTNTITKLTNVSINIGSNLSHLILHSKRGORGVOIDGIT 539  
 Db 480 KTLTNTTISNLYKNWMTNTITKLTNVSINIGSNLSHLILHSKRGORGVOIDGIT 539  
 Qy 540 ----SKGNLTYISGGWVDVHKNTLD-QGFENLITA-ASVAFEGCNKAKDAANAKIVAQ 593  
 Db 540 ----SKGNLTYISGGWVDVHKNTLD-QGFENLITA-ASVAFEGCNKAKDAANAKIVAQ 593  
 Qy 594 GTVTITGEGKDFRANNVSLNGTGGKLNLISSVNN---LTHNLSGTINTISNITNITQTRK 650  
 Db 594 GTVTITGEGKDFRANNVSLNGTGGKLNLISSVNN---LTHNLSGTINTISNITNITQTRK 650  
 Qy 591 GTTIT-SGNQKGRFRNNVSLNGTSGLOPPTTKRTNKYAITNKFEGTLNISKVNISMVLPK 649  
 Db 591 GTTIT-SGNQKGRFRNNVSLNGTSGLOPPTTKRTNKYAITNKFEGTLNISKVNISMVLPK 649

Qy 651 NTS-YWOTSHDSHNVVSAINLETGANFTF-IKYISSNSKGLTTOYRSSAGVNFVNGVM 708  
 Db 650 NESGYDKFKRTYWNLTSLNVSSEGEFNLITDSRSDSAGTLTQPYNLNGISF---NKDT 706  
 Qy 709 SFNLKGAQVNFKLPENNMNTSKLP-IRPLANTATGGGVFFDIYANISG---RGAE 764  
 Db 707 TFNVARNVNFEDIKAPIGINKYSSLVASFNGNLSVGGGVDFTLASSNVOTPGV 766  
 Qy 765 LKMEISINSGANFTLNSHVGRDDAFKINKDLTINATNSFSLRTKDDFDVGYARNAIN 824  
 Db 767 INSKYFNVSSTGSLRFKTSSTKTKTGFSEKDLTLNATGNITLLQVEGT--DGMTGKGI 824  
 Qy 825 STYNISILGNGVNTLGGQNSSSSITGNITIEKAAANTLLEANNAPNOQNTIRDRVILGSLV 884  
 Db 825 AKKNITFEGGNTFGRKAVTEIEGNVTINNANVTILGSDFDNHQ--KPLTIKKDVIN 882  
 Qy 885 NGSLSLTGENDIKGNLTISESATFKGTRDTLNTGNFTNNGTAEINITQGVKLGNTV 944  
 Db 883 SCNLTAGGNIVNLAGNLTVESNANFKAITNFTFNVGGLFDNKGNSNISAKGARFKDID 942  
 Qy 945 NDGDLNITTHAKRNORSIIGGDIINKKGSNLITDSNDAAEIOIGNISOKEGNLTISDK 1004  
 Db 943 NSKNLSITNSSSTYRTIISGNITNKNGLDNLITNEGSDTEMOIGGDVSKOEGNLTISDK 1002  
 Qy 1005 INITKOITTKGIDGEDSSDATSNANLTIKTKELKLFEDLSISGFNKAETAKDGRDLT 1064  
 Db 1003 INITKOITKAGVDGENSDATNANLTIKTKELKLTODLNISGFNKAETAKDGRDLT 1062  
 Qy 1065 IGNSDNGSGAEAKTVTFNVKDKSISADGHNVTLNSKVYKTSSSNGGRESNDTGLTI 1124  
 Db 1063 IGNTNSAD-GTNAKVTNFQVQDKSISADGHKVTLLHSKVETSGSNNTEDSDNNAGLTI 1121  
 Qy 1125 TAKNVNVDITSLTKVNITA-SEKVTITAGSTINATNGKASITTKT----- 1170  
 Db 1122 DAKNVTVNNITSHKAVSISATSGETTITGTINATTGNVEITAGTSGILGIESSSGS 1181  
 Qy 1171 -----GDISGTISGNT 1181  
 Db 1182 VTLTATEGALAVSNISGNTVTVTANSALTLTLAGSTIKGTESVTTSSQSGIGGTISGTT 1241  
 Qy 1182 VSVASVVDLTTSKSGIEAKSGEANVTSTATGTTIGTISGNTVNTANAGDLTVNGAEIN 1241  
 Db 1242 VEKATESLTQTSNISKATTEANVTSTATGTTIGTISGNTVNTANAGDLTVNGAEIN 1301  
 Qy 1242 ATEGAATLTGNTLTEAGSSITSTKGOVDLLAONGSTAGSINANVTNTGTLTTVA 1301  
 Db 1302 ATEGAATLTSSCKLTTEASSHTSAGQVNLQAQDSVAGSINANVTNTGTLTTVK 1361  
 Qy 1302 GSDIKATSGTLVINAKDAKNGDASGDSDEVNAVNASGSGSVTAATSSSVITGDLNTVN 1361  
 Db 1362 GSNINATSGTLVINAKDAKNGDASGDSDEVNAVNASGSGSVTAATSSSVITGDLNTVN 1421  
 Qy 1362 GLNIIISKDGRNTVRLKGEIEVKYIQPGVASVEEVEIAKRVLEKVKDLSDERETLAKLG 1421  
 Db 1422 GLNIIISKNGINTVLLKGVKIDVKYIQPGVASVEEVEIAKRVLEKVKDLSDERETLAKLG 1481  
 Qy 1422 VSAVREVEPNNTITVNTQNEFTTRPSSQVSIISSEKACFSGSGNAGVCTVNAODGQ 1476  
 Db 1482 VSAVREVEPNNTITVNTQNEFTTRPSSQVSIISSEKACFSGSGNAGVCTVNAODGQ 1536

## RESULT 12

AA001846  
 ID AAB01846 standard; Protein; 1536 AA.

XX AAB01846;

XX AC AAB01846;

XX DT 11-SEP-2000 (first entry)

XX Haemophilus influenzae strain 12 HmW1A protein, SEQ ID NO:67.

DE HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
 KW non-typeable Haemophilus influenzae; NTH1; non-encapsulated;



recombinant production; *Escherichia coli*; antibacterial; vaccine;  
human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
detection; diagnosis.

Haemophilus influenzae strain 12.

WO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

07-OCT-1998; 98US-0167568.

08-DEC-1998; 98US-0206942.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-303789/26.

N-PSDB; AAA52195.

Nucleic acid molecule for producing recombinant high molecular weight  
proteins of Haemophilus which are used as a vaccine to provide  
protection against Haemophilus induced diseases in humans -

Example 16; Fig 28A-Q; 307pp; English.

The invention relates to the recombinant production of Haemophilus  
influenzae high molecular weight (HMW) proteins in *Escherichia coli*. The  
expression construct used to effect recombinant expression comprises a  
promoter functional in *E. coli* (e.g., the T7 promoter) operably linked  
to a modified hmwABC operon from a non-typeable (non-encapsulated) H.  
influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene  
clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,  
hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins  
and the hmwB and hmwC genes encode accessory proteins which are  
responsible for post-translational processing and secretion of the HMW  
proteins. The modified hmwABC operon used in the expression construct of  
the invention contains an A gene modified such that it encodes only the  
mature HMW. The invention also discloses hmwA genes (AA52175-AA52198)  
strains Joyce, K1, K21, LDCD2, PMH1, 15 and 12. The nucleic acids and  
vectors are used for the production of recombinant H. influenzae HMW  
proteins which can be used as vaccines to mediate a humoral or  
cell-mediated immune response to provide protection against diseases in  
humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
pneumonia and tracheobronchitis). The HMW proteins are also useful as  
antigens in immunoassays for detecting antibodies against Haemophilus,  
HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
HMW proteins can be used to isolate and clone hmw genes from other  
non-typeable strains of Haemophilus via hybridisation reactions. The  
present sequence represents an HMW protein from a non-typeable strain of  
H. influenzae.

SQ Sequence 1536 AA;

Query Match 66.7%; Score 4943; DB 21; Length 1536;

Best Local Similarity 67.1%; Pred. No. 5.4e-227;

Matches 1043; Conservative 147; Mismatches 267; Indels 98; Gaps 20;

QY 1 MNKTYRLKSKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLSCVT 60

DB 1 MNKLYRLKSKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLSCVT 60

QY 61 SIPOSVLASGLQGVVHGATQVQDGNKTIIRNSVDIINWKQFNIDQNMVQFLOENN 120

DB 61 SIPOSVLASGLQGVVHGATQVQDGNKTIIRNSVDIINWKQFNIDQNMVQFLOENN 120

QY 121 NSAVFNRTSNQISQLKGLDSNGOVFLINPNGITIGKDALINTNGFTASTLDSINENIK 180

DB 121 NSAVFNRTSNQISQLKGLDSNGOVFLINPNGITIGKDALINTNGFTASTLDSINENIK 180

QY 181 ARNFTTEQTKD KALAEIVNHLITVKGDSVNLIGKVKNEGVI SVNGGSISLLAGOKIT 240  
DB 181 ARNFTTEQTKD KALAEIVNHLITVKGDSVNLIGKVKNEGVI SVNGGSISLLAGOKIT 240  
QY 241 ISDIINPTTISTAA PENEA VNLGDI FAKGGINVRAA TIRNOGKLSADSVSKDKSGNIV 300  
DB 241 ISDIINPTTISTAA PENEA VNLGDI FAKGGINVRAA TIRNOGKLSADSVSKDKSGNIV 300  
QY 301 LSAKEGEAEIGGVISAQNOQA KGLMTITGDKVTLKTGAVIDLSGEGGETYLGGBERGE 360  
DB 301 LSAKEGEAEIGGVISAQNOQA KGLMTITGDKVTLKTGAVIDLSGEGGETYLGGBERGE 360  
QY 361 GKNGIOLAKKTSLEKSTINVSCKEKGGAIVWGDIALIDGNINAQSGDIAKTGFVET 420  
DB 361 GKNGIOLAKKTSLEKSTINVSCKEKGGAIVWGDIALIDGNINAQSGDIAKTGFVET 420  
QY 421 SCHDLFIKDNAIVDAKEWLLDFDNVYSINAEDPLFNWNTGINDPEPTGTG-EASDPKKNSEL 479  
DB 421 SCHDLFIKDNAIVDAKEWLLDFDNVYSINAEDPLFNWNTGINDPEPTGTG-EASDPKKNSEL 479  
QY 480 KTTLTNTTISNYLKNAWMTNITASRKLTVNSSINIGSNHSLHLHSGQGGGVVIDGDIT 539  
DB 479 KTTLTNTTLESILKKGTFVNITANQRIYVNSSINL-SNGSLTLWSEGRSGGVVEINNDIT 537  
QY 540 ----SKGNLTIYSGWVDVHKNTILD-QGFLNITA-ASVAFEGGNKKAARDAAAKIVAQ 593  
DB 538 TGDDTRGANLTIYSGWVDVHKNTISLGAQGNINITAKQDIAFEKGSNQV-----ITGQ 590  
QY 594 GTVTITGCKDFRANVSLUNGTKGLNISSVNN--LTHNLSGTINISGNTINOTTRK 650  
DB 591 GTIT-SGNQKGRFRNVSLUNGTCGLOFTTKTNYAITNKFEGLTNISGKNISMVLPK 649  
QY 651 NTS-YWQTSHDSSHNVSAALNLTGANFTF-IKYISSNSKGLTQYRSSAGVNFNGVNGM 708  
DB 650 NESGYDKFKGRTYWNLTSLNSESCEFNLTIDSRGSDSAGTLTPQVNLNGISF---NKDT 706  
QY 709 SFNLKEGAKVNFKLKPNENMNTSKPLP-IRFLANTATGGGVDFDIYANHSG---RGAE 764  
DB 707 TFNVERNARVNDIKAPIGINKYSLNYSFNGNISVSGGSDVDTLLASSNVQTPGVV 766  
QY 765 LKMSINISNGANFTLNSHVRGDDAFKINKDLTINATSNFSLRQTKDDFYDGYARNAIN 824  
DB 767 INSKYFNVSTGSSLRFTKSGSTKTCFSEKDLTLANATGNTLLQVEGT--DGMIGKGLV 824  
QY 825 STYNISILGWNVTGGQNSSSITGNITIEKAANYTLEANNAPNOQNIIRDRVILGLSLV 884  
DB 825 AKKNITFEGGNITFGSRKAVTEIEGCVTINNANVTLLIGSDFDNHQ---KPLTIKKDVIIN 882  
QY 885 NGSLTUCENADIKGNLTISATFKGKTRDTLNTGNTNGTAETINLTQGVVKGWVT 944  
DB 883 SGNLTAGNIYVNIAGNLTVESNANFKAITNFTFNVGGLDFNKGNSNISIAKGARFKDID 942  
QY 945 NGDLNLTTHAKRNQRIIGGDIINKKGLSLNITDSNDAETIQIGNISQKGNLTISDDK 1004  
DB 943 NSKNLSITNSSSTYRTIISGNTNKNGLDNLITNEGSDTEMQIGGDSVQKGNLTISDDK 1002  
QY 1005 INITKQITIKKIGDCEBSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLT 1064  
DB 1003 INITKQITIKAGVDGSDSDATNNANLTIKTKELKLTQDLNISGFNKAETAKDGSALT 1062  
QY 1065 IGNSDNGSGAEAKTVTFNVNVDKSKISADGHNVTLNKS VKTSSSSNGGRESNDNDTGLTI 1124  
DB 1063 IGNTNSAD-GTNNAKVVTFNQVKDKSKISADGHKVTLSHKVETSGSNNTEDSDNNAGLTI 1121  
QY 1125 TAKNVEYKDTISLKTVNITA-SEKVTITTAGSTINAYNGKASITTKT----- 1170  
DB 1122 DAKNVTNNITSHKAVSISATSGEITTKGTTINATGTVNVEITAQIGSLIGGIESSSGS 1181  
QY 1171 -----GDISGTISGNT 1181  
DB 1182 VTLTATEGALAVSNISGNTVTVTANSALATTLAGSTIKGTESVTTSSQSGDIGTISGNT 1241  
QY 1182 VSVSATVDLTTKSGSKTEAKSGEANVT SATGCT ICGTISGNTVNTVNTANAGDLTVGNGAEIN 1241

Db 1242 VEVKATESLTQSNKIKATTGEANVTATGTTIGTISGNTVNVNANAGDLTVGNGAEIN 1301  
 QY 1242 ATEGAATLPAATNTLTTEAGSSITSTKGQVDLLAQAQSIAGSINAANYTLNLTGTLTTVA 1301  
 Db 1302 ATEGAATLTSSGKLTTEASSHTSAGKQVNLQAQSGSVAGSINAANYTLNLTGTLTTVK 1361  
 QY 1302 GSDIKATSTLVINAKDAKLDGASDSTEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1361  
 Db 1362 GSNINATSTLVINAKDAELNGAALNHTVNVNATNANGSGSVIATSSRVNITGDLTIN 1421  
 QY 1362 GLNIIISKGRNTVRLRGKEIEVKYIOPGVASVEEVEAKRVLEKVKDLSDEEREFLAKLG 1421  
 Db 1422 GLNIIISKNGINTVLLKGVKIDVKYIOPGIVASVDEVEAKRVLEKVKDLSDEEREALAKLG 1481  
 QY 1422 VSAVRFEVNNITVNTONEFTTRSSOVIIISGKACFSSGNGARVCTNVADGQ 1476  
 Db 1482 VSAVRFEVNNITVNTONEFTTRSSOVIIISGKACFSSGNGARVCTNVADGQ 1536

## RESULT 13

AAR41725  
 ID AAR41725 standard; Protein; 1536 AA.

AC AAR41725;

DT 26-APR-1994 (first entry)

XX High molecular weight protein 1 (HMW1).

XX HMW; high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae; gene cluster.

XX Haemophilus influenzae.

XX Key Location/Qualifiers

FT Misc-difference 668..677  
 FT /note= "Possibly incorrect sequence. Alternative  
 FT sequence for this region is LNVSEGEFN.  
 FT (see comments)."

XX WO9319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993; 93WO-US02166.

XX 16-MAR-1992; 92GB-0005704.

XX (BARE/) BARENKAMP S J.  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

XX N-PSDB; AAO49508.

XX High molecular weight surface proteins - of non-typeable

XX haemophilus which exhibit immunogenic properties

XX Claim 3; Figure 2/10; 100pp; English.

XX The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines. This sequence is claimed to be  
 CC the same as that given in AAR41723 (High molecular weight protein 1)  
 CC although it does differ slightly. (Repeated regions which are  
 CC possibly incorrect and occur in the corresponding nucleotide coding  
 CC sequence contribute to these differences).

XX Sequence 1536 AA;

SQ

Query Match 66.5%; Score 4924; DB 14; Length 1536;  
 Best Local Similarity 67.0%; Pred. No. 4.3e-226;  
 Matches 1042; Conservative 147; Mismatches 268; Indels 98; Gaps 20;

QY 1 MNKYIRLKFKRLNALVAVSELARGCDHSTKGESEKFAKMKVRLHAKLPISAMLLSLGVT 60  
 Db 1 MNKYIRLKFKRLNALVAVSELARGCDHSTKGESEKFAKMKVRLHAKLPISAMLLSLGVT 60  
 QY 61 SIPOSVLASGLQGMVYVHGATMQVDGNKTIIRNSVDALINWKFNDQNMWQVFLQENN 120  
 Db 61 SIPOSVLASGLQGMVYVHGATMQVDGNKTIIRNSVDALINWKFNDQNMWQVFLQENN 120  
 QY 121 NSAVFNRTVTSNOISQLKGLDSNGOVFLNPGITIGKDAINTNGFTASTLDISNENIK 180  
 Db 121 NSAVFNRTVTSNOISQLKGLDSNGOVFLNPGITIGKDAINTNGFTASTLDISNENIK 180  
 QY 181 ARNFTFEOTKDALAEIVNHGLITVKGDSVNLIGGKVKNEGIVSVNGSGISLLAGOKIT 240  
 Db 181 ARNFTFEOTKDALAEIVNHGLITVKGDSVNLIGGKVKNEGIVSVNGSGISLLAGOKIT 240  
 QY 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSKDSGNIV 300  
 Db 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSKDSGNIV 300  
 QY 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSKGEGETYLGDERGE 360  
 Db 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSKGEGETYLGDERGE 360  
 QY 361 KNGIQLAKTSLKSGSTINVSKEKGGFAIWDGIALIDGNIINAGSGSDIAKTGGFVET 420  
 Db 361 KNGIQLAKTSLKSGSTINVSKEKGGFAIWDGIALIDGNIINAGSGSDIAKTGGFVET 420  
 QY 421 SGHDLFIKDAIVDAKEMWLLDPNVNSINAEPLFNNTGINDDEFPTGTG-EASDPEKKNSEL 479  
 Db 421 SGHDLFIKDAIVDAKEMWLLDPNVNSINAEPLFNNTGINDDEFPTGTG-EASDPEKKNSEL 479  
 QY 480 KTTLTNTTISYLNKAMTNITASRKLTNVNSINTGNSHLILHKSQGGQGVQIDGIT 539  
 Db 480 KTTLTNTTISYLNKAMTNITASRKLTNVNSINTGNSHLILHKSQGGQGVQIDGIT 539  
 QY 540 ----SKGGLTIYSGGVWDVHKNIITLD-OGFLNITA-ASVAFEGGNNKARDAANAKIVAO 593  
 Db 540 ----SKGGLTIYSGGVWDVHKNIITLD-OGFLNITA-ASVAFEGGNNKARDAANAKIVAO 593  
 QY 594 GTVTTITGEGKDFRANNVSLNGTGKGLNIISVNN--LTHNLSGTINISGNTITINQTRK 650  
 Db 594 GTVTTITGEGKDFRANNVSLNGTGKGLNIISVNN--LTHNLSGTINISGNTITINQTRK 650  
 QY 651 NTS-YWOTSHDSHNV-SALNLETGANFTFIKYISSNKGTLTOYRSSAGVNFVNGVNM 708  
 Db 651 NTS-YWOTSHDSHNV-SALNLETGANFTFIKYISSNKGTLTOYRSSAGVNFVNGVNM 708  
 QY 709 SPNLKEGAKVNFKLKPNENWNTSKPLP-LRFLANITATGGSVFFDIYANHS--RGAE 764  
 Db 709 SPNLKEGAKVNFKLKPNENWNTSKPLP-LRFLANITATGGSVFFDIYANHS--RGAE 764  
 QY 765 LKMEINISNGANFTLNSHVRGDDAFKINKDLTINATNSFSLRQKDDFYDGYARNAIN 824  
 Db 765 LKMEINISNGANFTLNSHVRGDDAFKINKDLTINATNSFSLRQKDDFYDGYARNAIN 824  
 QY 825 STYNISILGGNVTLGGQNSSSSITGNITIEKAAVNTLEANNAPNOQNIIDRVIKGLSLV 884  
 Db 825 STYNISILGGNVTLGGQNSSSSITGNITIEKAAVNTLEANNAPNOQNIIDRVIKGLSLV 884  
 QY 885 NGSLISITGENADIKGNLTITSESATFKGKTRDLTNITNGFTNNGTAEINTQGVKLGNYT 944  
 Db 885 NGSLISITGENADIKGNLTITSESATFKGKTRDLTNITNGFTNNGTAEINTQGVKLGNYT 944  
 QY 945 NQDGLNITTHAKRNORSIIIGDIIINKKGLNITDSSNDAEIQIGGNISQKEGNLTSSDK 1004  
 Db 945 NQDGLNITTHAKRNORSIIIGDIIINKKGLNITDSSNDAEIQIGGNISQKEGNLTSSDK 1004

QY 1005 INITKQITIKKIDGEDSSDSDATSNANLTIKTKELKLTDLTSGFNKAEITAKGDRDLT 1064  
 Db 1003 INITKQITIRAGVDGSDSDATSNANLTIKTKELKLTDLTSGFNKAEITAKGDSDLT 1062  
 QY 1065 IGNSDNGSAGAEKVTFFNNVDSKISADGHNVTLSKYKVTSSNGSGRESNDNDTGLRI 1124  
 Db 1063 IGNTNSAD-GTNNAKVTFFNOVKDSKISADGKHVTLHSHKVTSSNNTEDSDNNAGLTI 1121  
 QY 1125 TAKNEVNDKITSIKVTNITA-SEKVTYTAGSTINATNGKASITTKT----- 1170  
 Db 1122 DAKNVTNNITSHKAVSISATSGEITTKTGTINATTGNVEITAGTSGILGIESSSGS 1181  
 QY 1171 -----GDISGTISGNT 1181  
 Db 1182 VTLTATEGALAVSNISGNTVTVTANSALTLTLAGSTIKETESVTSSQSGDGGTISGNT 1241  
 QY 1182 VVSATVADLTTRSGKIEAKSGEANTVSATGTIGGTISGNTVNVNANAGDLTVNGAEIN 1241  
 Db 1242 VEKATESLUTQSNSKIKATTEGEANTVSATGTIGGTISGNTVNVNANAGDLTVNGAEIN 1301  
 QY 1242 ATEGAATLTATGNTLTTEAGSSITSTKGVLDLLAONGSTAGSINAAVNTLTGTTLTVA 1301  
 Db 1302 ATEGAATLTSSGKLTTEASHITSAKGVNLSAQDGSVAGSINAAVNTLTGTTLTVK 1361  
 QY 1302 GSDIKATSLVINAKDAGLNGDSCDSTEYNAVNASGSGSVTAATSSSVNITGDLNTVN 1361  
 Db 1362 GSNINATSGTLVINAKDAELNAGLNTVNVNATNANGSGSVIATSSRVNITGDLTIN 1421  
 QY 1362 GLNIISKGRNVRGRGIEVKYIOPGVASVEEVEIAKRVLEKVDLSDEERETLAKLG 1421  
 Db 1422 GLNIISKGINVLKGVKIDVKYIOPGVASVDEVEIAKRVLEKVDLSDEERETLAKLG 1481  
 QY 1422 VSAVREVPNNITVTNTONEFTTRPSSOVITISEGKACFPSSGNGARVCTNVAADQO 1476  
 Db 1482 VSAVREVPNNITVTNTONEFTTRPSSOVITISEGKACFPSSGNGARVCTNVAADQO 1536

RESULT 14  
 AAW30291  
 ID AAW30291 standard; Protein; 1598 AA.  
 AC AAW30291;  
 XX  
 DT 14-APR-1998 (first entry)  
 XX  
 DE Non-typeable Haemophilus high mol.wt. surface protein HMW3.  
 KW Non-typeable Haemophilus; high molecular weight surface protein;  
 KW HMW3; immunogen; vaccine; otitis media.  
 XX  
 OS Haemophilus influenzae strain 5.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 113 /note= "encoded by GTG"  
 FT Misc-difference 864 /note= "encoded by TGT"  
 XX  
 PN WO9736914-A1.  
 XX  
 PD 08-OCT-1997.  
 XX  
 PF 01-APR-1997; 97WO-US04707.  
 XX  
 PR 01-APR-1996; 96US-0617697.  
 XX  
 PA (BARE/) BARENKAMP S J.  
 XX  
 PI Barenkamp SJ;  
 XX  
 DR WPI: 1997-503038/46.  
 DR N-PSDB: AAW30291.  
 XX

High molecular weight proteins of non-typeable Haemophilus influenzae - useful for vaccine production

Claim 1; Page 93-97; 183pp; English.

This protein comprises the high molecular weight surface protein HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that has the immunological ability to protect against disease caused by a non-typeable Haemophilus strain and is characterised by at least one surface-exposed B-cell epitope that is recognised by monoclonal antibody AD6. The HMW3 amino acid sequence was deduced from an isolated hmw3 gene (see AAT90992). HMW1 (see AAW30293), HMW2 (see AAW30294) and HMW4 (see AAW30292) have also been identified. A conjugate comprising HMW3 linked to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150 amino acids corresponding to at least protective epitope of HMW3 are also claimed. HMW proteins, conjugates and peptides can be used in vaccines, as immunogens for preparation of antibodies and as antigens for detection of these antibodies.

Sequence 1598 AA;

Query Match 64.1%; Score 4748.5; DB 18; Length 1598;  
 Best Local Similarity 62.5%; Pred. No. 1e-217;  
 Matches 1019; Conservative 143; Mismatches 281; Indels 187; Gaps 25;

QY 1 MNKIYRLKFKSRNLNVAVSELARGCDHSTKGESEKPMKVRHLAKPLSALLSLGVT 60  
 Db 1 MNKIYRLKFKSRNLNVAVSELARGCDHSTKGESEKPMKVRHLAKPLSALLSLGMA 60  
 QY 61 SIPOSVLASLQGVHVGATMVDGNTKIIRNSVDAILNWKQFNIDQNMVQFQENN 120  
 Db 61 SIPOSVLASLQGVHVGATMVDGNTKIIRNSVDAILNWKQFNIDQNMVQFQENN 120  
 QY 121 NSAVENRVTNSQLSGLDLSNGQVFLNPGITTKGDAIINTNGFTASTLDSINENIK 180  
 Db 121 NSAVENRVTNSQLSGLDLSNGQVFLNPGITTKGDAIINTNGFTASTLDSINENIK 180  
 QY 181 ARNFTFQTKDALKAEIVNHLITVGVKDSVNLIGKVKNEGVSIVNGSISLLAGOKIT 240  
 Db 181 ARNFTFQTKDALKAEIVNHLITVGVKDSVNLIGKVKNEGVSIVNGSISLLAGOKIT 240  
 QY 241 ISDIINPTITYSTAAPENAEVNLGDIKAGGNINVRAATIRNOGLSADSVSKDSGNIV 300  
 Db 241 ISDIINPTITYSTAAPENAEVNLGDIKAGGNINVRAATIRNOGLSADSVSKDSGNIV 300  
 QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGETYLGDERGE 360  
 Db 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGETYLGDERGE 360  
 QY 361 GKNGIOLAKKTSLEKGSTINVSKEKGGFAIVMGDIALIDGNINAQGGDIKAGTGFVET 420  
 Db 361 GKNGIOLAKKTSLEKGSTINVSKEKGGFAIVMGDIALIDGNINAQGGDIKAGTGFVET 420  
 QY 421 SGHDLFIKDNAIVDAKWLDFDNVSINAEOPLFNNTGIDNDEFPTGTGEADSP--KKNSE 478  
 Db 421 SGHDLFIKDNAIVDAKWLDFDNVSINAEOPLFNNTGIDNDEFPTGTGEADSP--KKNSE 478  
 QY 479 LKTLTNTTISNLYLNAWNTMTASRKLTVNSISIGNSHLILHSKGGGGVQIDGDI 538  
 Db 479 LKTLTNTTISNLYLNAWNTMTASRKLTVNSISIGNSHLILHSKGGGGVQIDGDI 538  
 QY 539 TSKGNLTIYSGGVVDVHKNTLDOGFNITA--ASVAFE--GGNNKARDANAANAKIYAQ 594  
 Db 539 TSKGNLTIYSGGVVDVHKNTLDOGFNITA--ASVAFE--GGNNKARDANAANAKIYAQ 594  
 QY 595 TVTITGEGKDFRANVSLNGTGGKGLNISS-----VNNLTNLSGTINISNINI 646  
 Db 595 TVTITGEGKDFRANVSLNGTGGKGLNISS-----VNNLTNLSGTINISNINI 646  
 QY 647 TTRKNTSWQTSKSHNVNVSALNLETGANFTF--IKYISSNSKGLTQYRSAGVNFNGVN 705  
 Db 647 TTRKNTSWQTSKSHNVNVSALNLETGANFTF--IKYISSNSKGLTQYRSAGVNFNGVN 705  
 QY 651 KAPKVSWFYRDKGRYWNVTTLNVTSGSKFNLSIDTSGSGTGPSIRNAELNGITFN--- 707  
 Db 651 KAPKVSWFYRDKGRYWNVTTLNVTSGSKFNLSIDTSGSGTGPSIRNAELNGITFN--- 707

```

QY 706 GNMSENLKEGAKVNEFLKPNENMTSKPLPIRFLAN-----ITATGGSVFFEDIYANH 758
Db 708 -KATFNIAOGSTANFSIK-----ASIMPEKSNAYALFNEDISVSGGSLNFKLNASS 759
QY 759 SG---RGAELKMEINISGANTFLNASHVRGDDAFKINKDLTINATNSNFSLRQTKDDFY 815
Db 760 SNIOTPGVVIKSONFNVSGSTLNLKAEGSTETAFSIENDLNLNAITGGNITIROVEGT-- 817
QY 816 DGTARNAINSTYISILGQNVTLGGQSSSTIGNLTIEKAANVTLE-ANNAPOO--NI 872
Db 818 DSRVKGVAAKNITFGKNITFGSKATYIEIKGNVTINKNTNATLARGANFAENKSPINI 877
QY 873 RDRVILGSLVNGSLTGENADIKGNLTISATFKGKTRDTLNTITNGTNGTAEIN 932
Db 878 -----AGNVINGNLTLTAGSIINIAGNLTVSKANLQALNTYFNAGSFDNNGASNIS 931
QY 933 ITQGVVVLGNVNDGLNITHAKRNQRSIIGDIIKNGSLNLTDSNDAEIOIGNIS 992
Db 932 IARGAKFKDINTSSNLNITNSDTYRTIIGKINISKSGDLNIDKSDAEIOIGNIS 991
QY 993 QKEGNLTISDKINIKQITIKKIDGEDSSDATSNANLTIKTKELKLTEDLSISGPNK 1052
Db 992 QKEGNLTISDKVNIITQITIKAGVEGGRSDSEAEANLTIQTKELKLAGDLNISGPNK 1051
QY 1053 AEITAKDGRDLTIGNSNDGSGAEAKTVTFNNYKDKISADGHNVTLSNKSVKTTSSNGR 1112
Db 1052 AEITAKNGSOLTIGNSGGN--ADAKKVTDFKVDKSKISTDGHVTLNSEVKT--SNGSS 1107
QY 1113 ESNSDNDTGLTITAKNVENKDIITSLKTVNI-----1143
Db 1108 NAGNDSTGLTISAKDVTNNVNTSHKTIINISAAAGNVTTKEGTTINATGTSVEVTAQNG 1167
QY 1144 -----TASGKVTNTAGSTINATNGKASITTKTGDISG-----TISG 1179
Db 1168 TIKGNITSONVTVTATENLVNTENAVINATSGTVNISTKTGDIKGGIESTSGNVNITASG 1227
QY 1180 NTVSVS-----ATVD---LTPKSGSKTEAKSGEANVTSATG-----1212
Db 1228 NTLKSNITGODVTVTADAGALTTAGSTISATGNANITTKTDINGKVESSSGVTLV 1287
QY 1213 -----TIG-----1212
Db 1288 ATGATLAVNIGSNVTIITADSGKLTSTVGSTINGTNSVTTSQSOGDIEGTISGNTVNT 1347
QY 1227 ANAGDLTVNGAEINTEGAATLTATGNTLTTEAGSSITSTKGQVDDLAAONGSIAGSINA 1286
Db 1348 ASTGDLTIGNSAKVEAKNGAATLTAESGKLTQTGSSITSSNGOTTITAKDSSITAGNINA 1407
QY 1287 ANVTLTNTGTLTTVAGSDIKATSTGLTVINAKDAKLNGDASGDSSTEVNAVNASGSGVTAA 1346
Db 1408 ANVTLTNTGTLTGTGDSKINATSTGLTINAKDAKLDAAGSDRTVNAVATNAGSGNVTAK 1467
QY 1347 TSSSVNITGDLTNGNLNIIISKDGRNTVRLRGKEIEVKYIQPGVASVEEVEIARVLEKV 1406
Db 1468 TSSSVNITGDLTNGNLNIIISNGRNTVRLRGKEIDVKYIQPGVASVEEVEIARVLEKV 1527
QY 1407 KDLSEERETLAKLGVSAREVEPNNTITVNTQNEFTTRPSSQVIIIEGKACFSSNGAR 1466
Db 1528 KDLSEERETLAKLGVSAREVEPNNAITVNTQNEFTTKPSSQVTIIEGKACFSSNGAR 1587
QY 1467 VCTNVADPDQ 1476
Db 1588 VCTNVADPDQ 1597

```

```

RESULT 15
ID AAR41731
XX AAR41731 standard; Protein; 1338 AA.
AC AAR41731;
XX AAR41731;
DT 26-APR-1994 (first entry)

```

```

XX High molecular weight protein 3 (HMW3).
DE HMW: high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX Haemophilus influenzae.
PN WO9319090-A.
PD 30-SEP-1993.
PF 16-MAR-1993; 93WO-US02166.
PR 16-MAR-1992; 92GB-0005704.
XX (BARE/) BARENKAMP S J.
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI Barenkamp SJ;
XX WPI: 1993-320683/40.
DR N-PSDB; AQA49510.
XX High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
XX Claim 5; Figure 10; 100pp; English.
XX The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX Sequence 1338 AA;
QY Query Match 47.6%; Score 3525.5; DB 14; Length 1338;
Best Local Similarity 56.5%; Pred. No. 1.2e-159;
Matches 774; Conservative 136; Mismatches 273; Indels 187; Gaps 25;
QY 261 VNLGDIKFAKGGNINVAATIRNOGKLSADSVSKDKSGNIVLSAKGEAEIGGVISAQNOO 320
Db 1 INLGDIFAKGGNINVAATIRNKKLSADSVSKDKSGNIVLSAKGEAEIGGVISAQNOO 60
QY 321 AKGCKLMTGDKVTLTKTGAVIDLSGKEGGETYLGGERGEGNGIOLAKKTSLEKGSTIN 380
Db 61 AKGCKLMTGDKVTLTKTGAVIDLSGKEGGETYLGGERGEGNGIOLAKKTSLEKGSTIN 120
QY 381 VSGKEKGGPATVWGDIALIDGNINAQSGDIATGFGVETSGHDLFIKDAIVDAKEWLL 440
Db 121 VSGKEKGGRAIWDGIALIDGNINAQSK-DIATGFGVETSGHYLSIDDAIVKTKWELL 179
QY 441 DFNVSINAEPLFNNTGINDPEPTGTGEASDP--KKNSLKTTLTNTISNLYKNATWM 498
Db 180 DPNVTIEA--PSASRVELGADRNHSAEVIKVTLLKNNFTSLTTLTNTISNLLSAHV 237
QY 499 NITASRKLTVNSSINTGSHLILHSKQGRGGVQIDGITSKGNGNTIYSGWVDVHK 558
Db 238 NITARRKLTVNSSISIERGSHLILHSEGGQGVQDKDITSEGNLTITISGWDVHK 297
QY 559 ITLDQGFNLITA--ASVAFE--GGNNKARDAANAKIAVAGCTVTITGEGKDFRANNYS 614
Db 298 ITLGSGLFNITTKEGDIAFDKSGRN-----NLITTAOGTIT-SGNSNGFRFNYS 350
QY 615 TGKGLNISS-----VNNLTHNLSGTINISGNTINOTTRKNTSYWOTSHDSHWYS 666
Db 351 LGGKLSPTDSREDGRRTKGNISNKFDTGLNISGTVDSMKAPKVSFWYDKGRTYVNT 410
QY 667 ALNLETGANETFE-IKYISSNSKGLTTOYRSSAGVNFVNGVNGNMFNLKEGAKVNEFLKPN 725
Db 411 TLNVTSGSKFNLSIDTSGSGSTGPSIRNAELINGITFN-----KATFNIAOGSTANFSIK-- 464

```



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:24:19 ; Search time 18 seconds  
(without alignments)  
2414.314 Million cell updates/sec

Title: US-10-092-880-4

Perfect score: 7407

Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7407	100.0	1477	1	US-08-038-682-4 Sequence 4, Appl
2	7407	100.0	1477	1	US-08-302-832-4 Sequence 4, Appl
3	7407	100.0	1477	2	US-08-530-198-4 Sequence 4, Appl
4	7407	100.0	1477	2	US-08-469-880-4 Sequence 4, Appl
5	7407	100.0	1477	2	US-08-728-470-4 Sequence 4, Appl
6	7407	100.0	1477	2	US-08-617-697-4 Sequence 4, Appl
7	7407	100.0	1477	4	US-08-719-641-4 Sequence 4, Appl
8	7261	98.0	1477	4	US-09-206-942-71 Sequence 71, Appl
9	5282	71.3	1600	2	US-08-617-697-10 Sequence 10, Appl
10	5164	69.7	1036	4	US-09-206-942-73 Sequence 73, Appl
11	4967	67.1	1536	1	US-08-038-682-2 Sequence 2, Appl
12	4967	67.1	1536	1	US-08-302-832-2 Sequence 2, Appl
13	4967	67.1	1536	2	US-08-530-198-2 Sequence 2, Appl
14	4967	67.1	1536	2	US-08-469-880-2 Sequence 2, Appl
15	4967	67.1	1536	2	US-08-728-470-2 Sequence 2, Appl
16	4967	67.1	1536	2	US-08-617-697-2 Sequence 2, Appl
17	4967	67.1	1536	4	US-08-719-641-2 Sequence 2, Appl
18	4956	66.9	1529	2	US-08-728-470-10 Sequence 10, Appl
19	4956	66.9	1536	4	US-08-719-641-10 Sequence 10, Appl
20	4943	66.7	1536	4	US-09-206-942-67 Sequence 67, Appl
21	4751.5	64.1	1599	2	US-08-617-697-9 Sequence 9, Appl
22	3525.5	47.6	1338	2	US-08-728-470-9 Sequence 9, Appl
23	3525.5	47.6	1338	4	US-08-719-641-9 Sequence 9, Appl
24	3396.5	45.9	963	4	US-09-206-942-51 Sequence 51, Appl
25	3365.5	45.4	977	4	US-09-206-942-53 Sequence 53, Appl
26	3260	44.0	998	4	US-09-206-942-59 Sequence 59, Appl
27	3256.5	44.0	1079	4	US-09-206-942-47 Sequence 47, Appl

28 3229 43.6 992 4 US-09-206-942-61 Sequence 61, Appl  
29 3225.5 43.5 1073 4 US-09-206-942-49 Sequence 49, Appl  
30 3184 43.0 975 4 US-09-206-942-30 Sequence 30, Appl  
31 3153 42.6 969 4 US-09-206-942-32 Sequence 32, Appl  
32 2970.5 40.1 1011 4 US-09-206-942-39 Sequence 39, Appl  
33 2939.5 39.7 1005 4 US-09-206-942-41 Sequence 41, Appl  
34 2864 38.7 1010 4 US-09-206-942-55 Sequence 55, Appl  
35 2833 38.2 1004 4 US-09-206-942-57 Sequence 57, Appl  
36 2763 37.3 1095 4 US-09-206-942-69 Sequence 69, Appl  
37 2741 37.0 1101 4 US-09-206-942-43 Sequence 43, Appl  
38 2710 36.6 1095 4 US-09-206-942-45 Sequence 45, Appl  
39 2678.5 36.2 1226 4 US-09-206-942-26 Sequence 26, Appl  
40 2647.5 35.7 1220 4 US-09-206-942-28 Sequence 28, Appl  
41 2609.5 35.2 1180 4 US-09-206-942-65 Sequence 65, Appl  
42 2609.5 35.2 1188 4 US-09-206-942-63 Sequence 63, Appl  
43 2584.5 34.9 1228 4 US-09-206-942-34 Sequence 34, Appl  
44 2553.5 34.5 1222 4 US-09-206-942-37 Sequence 37, Appl  
45 2066.5 27.9 915 4 US-09-206-942-35 Sequence 35, Appl

## ALIGNMENTS

RESULT 1  
US-08-038-682-4  
; Sequence 4, Application US/08038682  
; Patent No. 5549897  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/038,682  
; FILING DATE: 16-MAR-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-293  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1477 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-038-682-4

Query Match 100.0%; Score 7407; DB 1; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPAKMKVRLHAKPLSAMLSLQVT 60  
Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPAKMKVRLHAKPLSAMLSLQVT 60



Qy	61	SIPOSVLASGLQGMVYVHGTTATQVODGNKTIIRNSVDAILNKKQFNIDQNMVQFLQENN	120
Db	61	SIPOSVLASGLQGMVYVHGTTATQVODGNKTIIRNSVDAILNKKQFNIDQNMVQFLQENN	120
Qy	121	NSAVFNRYTSNQISOLKGIILDSNGQVFLINPNGITIGKDAIINTNGFTASTLDISNENIK	180
Db	121	NSAVFNRYTSNQISOLKGIILDSNGQVFLINPNGITIGKDAIINTNGFTASTLDISNENIK	180
Qy	181	ARNFTFQOTKDALAEIVNHGLITTVCKDGSVNLIGKVKYKNEGVI SVNGGSIISLLAQKIT	240
Db	181	ARNFTFQOTKDALAEIVNHGLITTVCKDGSVNLIGKVKYKNEGVI SVNGGSIISLLAQKIT	240
Qy	241	ISDIINPTIYISAAPENEAVNLGDIFAKGGNINVRAATIRNOGKLSADSVSKDKSGNIV	300
Db	241	ISDIINPTIYISAAPENEAVNLGDIFAKGGNINVRAATIRNOGKLSADSVSKDKSGNIV	300
Qy	301	LSAKEGEAEITGGVISQAONOAKGKLMITGDKVTLTKTGAVIDLSKKEGETHYLGGDERGE	360
Db	301	LSAKEGEAEITGGVISQAONOAKGKLMITGDKVTLTKTGAVIDLSKKEGETHYLGGDERGE	360
Qy	361	GKNGIQIOLAKKTSLEKGSTINVSKEKGGGPAIVMGDIALIDGNIINAQGSDDIATKGGFVET	420
Db	361	GKNGIQIOLAKKTSLEKGSTINVSKEKGGGPAIVMGDIALIDGNIINAQGSDDIATKGGFVET	420
Qy	421	SGHDLFITKDNAYIDAKEWILLDFDNVINSINAEADPLFNNTGINDEFPTGTCEASDPKKNSELK	480
Db	421	SGHDLFITKDNAYIDAKEWILLDFDNVINSINAEADPLFNNTGINDEFPTGTCEASDPKKNSELK	480
Qy	481	TTLTNTTISNYLKNAMTMITASRKLTVNSSINIGSNHLLIHSKGQRGGVQIDGDIITS	540
Db	481	TTLTNTTISNYLKNAMTMITASRKLTVNSSINIGSNHLLIHSKGQRGGVQIDGDIITS	540
Qy	541	KGNLTIYISGGWYDVHKNTITLQGFPLINTAASVAPEGGNNKARDAANAKI VAQGTVTITG	600
Db	541	KGNLTIYISGGWYDVHKNTITLQGFPLINTAASVAPEGGNNKARDAANAKI VAQGTVTITG	600
Qy	601	EGKDFRANVYSLNGTCKGLNIISVNNLTHNLSGTINISGNITINOTTRKNTSVQWTSHD	660
Db	601	EGKDFRANVYSLNGTCKGLNIISVNNLTHNLSGTINISGNITINOTTRKNTSVQWTSHD	660
Qy	661	SHNVYSALNLETGANFTFTKYIISNSKGLITQYRSSAGVNFNGVNGNMSFNLKEGAKVNF	720
Db	661	SHNVYSALNLETGANFTFTKYIISNSKGLITQYRSSAGVNFNGVNGNMSFNLKEGAKVNF	720
Qy	721	KLKPNENMTSKPLIRFLANITATYGGGSVFEDIYANHSRGAEELKWEISINSGANFTL	780
Db	721	KLKPNENMTSKPLIRFLANITATYGGGSVFEDIYANHSRGAEELKWEISINSGANFTL	780
Qy	781	NSHVRGDDAFKINKOLITINATNSFNSLRQTKDDFYDGYARNAINSTYNISILGNNVTLGG	840
Db	781	NSHVRGDDAFKINKOLITINATNSFNSLRQTKDDFYDGYARNAINSTYNISILGNNVTLGG	840
Qy	841	QNSSSSITIGNITIEKAANYTLEANNAPNOQNIORVKIKGLSLLVNGSISLTGENADIKGN	900
Db	841	QNSSSSITIGNITIEKAANYTLEANNAPNOQNIORVKIKGLSLLVNGSISLTGENADIKGN	900
Qy	901	LTISATESKGTRODTLNTIGNFTNNGTAETINITQGVVKLGNTNDGDLNITTHAKRNQR	960
Db	901	LTISATESKGTRODTLNTIGNFTNNGTAETINITQGVVKLGNTNDGDLNITTHAKRNQR	960
Qy	961	SIIGGDIINKKSLNITDSNDDAEIQIGNTISQEGNLTITSSDKINITKOITIKKGIDGE	1020
Db	961	SIIGGDIINKKSLNITDSNDDAEIQIGNTISQEGNLTITSSDKINITKOITIKKGIDGE	1020
Qy	1021	DSSSDATSNANLTIKTKEKLKLTEDLISGFNKAETAKDGRDLTIGNSDNGNSGAFAKTIV	1080
Db	1021	DSSSDATSNANLTIKTKEKLKLTEDLISGFNKAETAKDGRDLTIGNSDNGNSGAFAKTIV	1080
Qy	1081	TFNNVYKDKISADGHNVITLNSKVYTSSSNGGRESNDOTGLTITAKNVEVNKDITSLKT	1140
Db	1081	TFNNVYKDKISADGHNVITLNSKVYTSSSNGGRESNDOTGLTITAKNVEVNKDITSLKT	1140

Qy	1141	UNITASEKVTTTAGSTINATNGKASITTTKGTGDISGPHISGNTVSVSATVDLTTLTKSGSKIEA	1200
Db	1141	UNITASEKVTTTAGSTINATNGKASITTTKGTGDISGPHISGNTVSVSATVDLTTLTKSGSKIEA	1200
Qy	1201	KSGEANVTSATCTIGTIGTSIGNTVNVTANAGDLTVNGCAEINATEGAATLTATGNTLTTEA	1260
Db	1201	KSGEANVTSATCTIGTIGTSIGNTVNVTANAGDLTVNGCAEINATEGAATLTATGNTLTTEA	1260
Qy	1261	GSSITSTKGVDLLAQNAGSIAGSINAAVNTLNTTGTLLTTVAGSDTKATSGTLVINAKDAK	1320
Db	1261	GSSITSTKGVDLLAQNAGSIAGSINAAVNTLNTTGTLLTTVAGSDTKATSGTLVINAKDAK	1320
Qy	1321	LNGDASGDSTEYNVNASGSGSVTAATSSSVNITGDLNTVNGLIISKDGRTVRLRGKE	1380
Db	1321	LNGDASGDSTEYNVNASGSGSVTAATSSSVNITGDLNTVNGLIISKDGRTVRLRGKE	1380
Qy	1381	IEVKYIQPGVASVEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVTNTON	1440
Db	1381	IEVKYIQPGVASVEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVTNTON	1440
Qy	1441	EFTTRPSSQVIAISEGKACFSSGNGARVCTNNVADGQP	1477
Db	1441	EFTTRPSSQVIAISEGKACFSSGNGARVCTNNVADGQP	1477

```

RESULT 2
US-08-302-832-4
; Sequence 4, Application US/08302832
; Patent No. 560398
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PS-08-302-832-4

```

Query Match 100.0%; Score 7407; DB 1: Length 1477;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MNKIYRLKSKRLNALVAVSELARCDHSTKSGSEKPARMKVRHIALKPLSAMLISLGV	60
Db	1	MNKIYRLKSKRLNALVAVSELARCDHSTKSGSEKPARMKVRHIALKPLSAMLISLGV	60
Qy	61	SIPQSVLASGLQCMQVVGHTATMOVDGKNTIIRNSVDALINWKQFNIDQNMVQFLOENN	120
Db	61	SIPQSVLASGLQCMQVVGHTATMOVDGKNTIIRNSVDALINWKQFNIDQNMVQFLOENN	120
Qy	121	NSAVFNRTVSNQISOLKGLDSNGOVFLINPNIGITIGKDAIINTNGFTASTLDISNENIK	180
Db	121	NSAVFNRTVSNQISOLKGLDSNGOVFLINPNIGITIGKDAIINTNGFTASTLDISNENIK	180
Qy	181	ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVI SVNGGSI SLLAGQKIT	240
Db	181	ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVI SVNGGSI SLLAGQKIT	240
Qy	241	ISDIINPTITYSTAAPENAVNLGDIKAGGNINVPAAIIRNOGKLSADSVSKDKSGNIV	300
Db	241	ISDIINPTITYSTAAPENAVNLGDIKAGGNINVPAAIIRNOGKLSADSVSKDKSGNIV	300
Qy	301	LSAKEGEAEITGGVISAQNOQAQGGKMLITGDKVTLKTGAVIDL SKEGGETV LGGDERGE	360
Db	301	LSAKEGEAEITGGVISAQNOQAQGGKMLITGDKVTLKTGAVIDL SKEGGETV LGGDERGE	360
Qy	361	GKNGIOLAKKTSLEKSTINVSKEKGGPAI VVGDIALIDGNITNAQGGSDIAKTGGFVET	420
Db	361	GKNGIOLAKKTSLEKSTINVSKEKGGPAI VVGDIALIDGNITNAQGGSDIAKTGGFVET	420
Qy	421	SGHDLFTKDAI VDAKEWILDFDNVSI NAEDPLFNNTGINDFPTGTGEASDPKKNSELK	480
Db	421	SGHDLFTKDAI VDAKEWILDFDNVSI NAEDPLFNNTGINDFPTGTGEASDPKKNSELK	480
Qy	481	TTLTNTTISNLYKNAMTNTASRKLT VNSSINIGSNHLLHSKGQGGVQIDGDI TS	540
Db	481	TTLTNTTISNLYKNAMTNTASRKLT VNSSINIGSNHLLHSKGQGGVQIDGDI TS	540
Qy	541	KGNLTYSGGWVDVHKNTLDOGFNLITTAASVAFEGGNKARDAANAKI VAGQTVITG	600
Db	541	KGNLTYSGGWVDVHKNTLDOGFNLITTAASVAFEGGNKARDAANAKI VAGQTVITG	600
Qy	601	EGKDFRANVSLNGTKGLMI ISSVNLTHNLSTINISGNIITINOTTRKNTSYWQTSHD	660
Db	601	EGKDFRANVSLNGTKGLMI ISSVNLTHNLSTINISGNIITINOTTRKNTSYWQTSHD	660
Qy	661	SHNVSALENGTGANFTFIKYSNSKGLTQYRSSAGVNFNGVNGMSEFNKEGAKVNF	720
Db	661	SHNVSALENGTGANFTFIKYSNSKGLTQYRSSAGVNFNGVNGMSEFNKEGAKVNF	720
Qy	721	KLKPNENMTSKPLPIRFLANITATGGGVFFDIYANHSGRGAE LKXSEINISNGANFTL	780
Db	721	KLKPNENMTSKPLPIRFLANITATGGGVFFDIYANHSGRGAE LKXSEINISNGANFTL	780
Qy	781	NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNTISILGCVNTLGG	840
Db	781	NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYNTISILGCVNTLGG	840
Qy	841	QNSSSITGNITIEKAANVLEANNAPNQINIRDRVFKLGLSVLNGSLTGENADIKGN	900
Db	841	QNSSSITGNITIEKAANVLEANNAPNQINIRDRVFKLGLSVLNGSLTGENADIKGN	900
Qy	901	LTISEATFKGKTROTTLNITGNFTNNGTAEINITQGVVVKLGNTVNDGDLNITTHAKRNOR	960
Db	901	LTISEATFKGKTROTTLNITGNFTNNGTAEINITQGVVVKLGNTVNDGDLNITTHAKRNOR	960
Qy	961	SIIGGDIINKGSLNITDSNNDAEIQGGNISQEGNLTISSDKINITKOITIKKGDGE	1020
Db	961	SIIGGDIINKGSLNITDSNNDAEIQGGNISQEGNLTISSDKINITKOITIKKGDGE	1020
Qy	1021	DSSSDATSNANLTIKTKEKLTEDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080
Db	1021	DSSSDATSNANLTIKTKEKLTEDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080

Db	1021	DSSSDATSNANLTIKTKEKLTEDLSISGPNKAEITAKDGRDLTIGNSDNGSGAEAKTV	1080
Qy	1081	TFNNVKDSKISAGCHNVTLSKVKTSNGSGRESNNDNDGLTITAKNNVEVNDITSLKT	1140
Db	1081	TFNNVKDSKISAGCHNVTLSKVKTSNGSGRESNNDNDGLTITAKNNVEVNDITSLKT	1140
Qy	1141	VNITASEKVTTAGSTINATNGKASITTKGDIISGTSIGNTVSVSATVDLTITKSGSKIEA	1200
Db	1141	VNITASEKVTTAGSTINATNGKASITTKGDIISGTSIGNTVSVSATVDLTITKSGSKIEA	1200
Qy	1201	KSGEANYTSATGTIGTISGNTVNTANAGDLTVNGAGAEINATEGAATLTATGNLTTEA	1260
Db	1201	KSGEANYTSATGTIGTISGNTVNTANAGDLTVNGAGAEINATEGAATLTATGNLTTEA	1260
Qy	1261	GSSITSTKGQVDDLLAONGSIAGSINAANVTLTNTTGTTLTVAGSDIKATSGTLVINAKDAK	1320
Db	1261	GSSITSTKGQVDDLLAONGSIAGSINAANVTLTNTTGTTLTVAGSDIKATSGTLVINAKDAK	1320
Qy	1321	LNGDASGSDTEVNANASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGENTVRLRGKE	1380
Db	1321	LNGDASGSDTEVNANASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGENTVRLRGKE	1380
Qy	1381	IEVKYIOPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLGVSAVRFEVPENNTITVNTON	1440
Db	1381	IEVKYIOPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLGVSAVRFEVPENNTITVNTON	1440
Qy	1441	EFTTRPSSQVLIISGKACFSNGGARVCTNVADDGQP	1477
Db	1441	EFTTRPSSQVLIISGKACFSNGGARVCTNVADDGQP	1477

RESULT 3

US-08-530-198-4  
; Sequence 4, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: JWB-1186  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1477 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-530-198-4

Query Match 100.0%; Score 7407; DB 2; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKIYRLKSKRLNALVAVSELARGDHSSTEGSEKPAKMKVPHLALPLSAMLISLGV 60  
DB 1 MNKIYRLKSKRLNALVAVSELARGDHSSTEGSEKPAKMKVPHLALPLSAMLISLGV 60

QY 61 SIPOSVLASGLOGMDVHVHTATMOVDGNKTIIRNSVDALINMKQFIDONEMVQFLQENN 120  
DB 61 SIPOSVLASGLOGMDVHVHTATMOVDGNKTIIRNSVDALINMKQFIDONEMVQFLQENN 120

QY 121 NSAVENRVTSMQISQKGLDLSNGQVFLNPNNGITIGKDAIINTNGFTASTLDISNENIK 180  
DB 121 NSAVENRVTSMQISQKGLDLSNGQVFLNPNNGITIGKDAIINTNGFTASTLDISNENIK 180

QY 181 ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSLLAGQKIT 240  
DB 181 ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSLLAGQKIT 240

QY 241 ISDIINPTIYTAAPENAVNLGDIKAGGKININRAATIRNOGKLSADSVSKDGSNIV 300  
DB 241 ISDIINPTIYTAAPENAVNLGDIKAGGKININRAATIRNOGKLSADSVSKDGSNIV 300

QY 301 LSAKEGEAEIGGVISAQNOQAKGGLMITGDKVTLTKTGAVIDLSKKEGGETYLGDERGE 360  
DB 301 LSAKEGEAEIGGVISAQNOQAKGGLMITGDKVTLTKTGAVIDLSKKEGGETYLGDERGE 360

QY 361 GRNGIQLAKKTSLEKGSTINSGKEKGGFAIVMGDIALIDGNIINAQSGDIAGTGGFVET 420  
DB 361 GRNGIQLAKKTSLEKGSTINSGKEKGGFAIVMGDIALIDGNIINAQSGDIAGTGGFVET 420

QY 421 SGHDLFIKDNATVDKAEWLLDPDNVSIINAEDPLFNNTGINDFPTGTGGEADPKKNSL 480  
DB 421 SGHDLFIKDNATVDKAEWLLDPDNVSIINAEDPLFNNTGINDFPTGTGGEADPKKNSL 480

QY 481 TTLTNTTISNVLKNAWTNITASRKLTVNSNINSGNSHLILHSKGQGGVQIDGDTIS 540  
DB 481 TTLTNTTISNVLKNAWTNITASRKLTVNSNINSGNSHLILHSKGQGGVQIDGDTIS 540

QY 541 KGNLTITSGGVVHKNITLDQGLFNITTAASVAFEGGNKARDAANAKIVAQGTVTITG 600  
DB 541 KGNLTITSGGVVHKNITLDQGLFNITTAASVAFEGGNKARDAANAKIVAQGTVTITG 600

QY 601 ECKDFRANVSLNGTKGLNIISSVNNLTHNLSTINISGNITINOTTRKNTSYWQTS 660  
DB 601 ECKDFRANVSLNGTKGLNIISSVNNLTHNLSTINISGNITINOTTRKNTSYWQTS 660

QY 661 SHWVNSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGMSFNLEKAGVNF 720  
DB 661 SHWVNSALNLETGANFTFIKYISSNSKGLTQYRSSAGVNFNGVNGMSFNLEKAGVNF 720

QY 721 KLPENNNTSKPLPIRFLANITATGGGVFFDIYANHSRGAEKLMSEINISNGANFTL 780  
DB 721 KLPENNNTSKPLPIRFLANITATGGGVFFDIYANHSRGAEKLMSEINISNGANFTL 780

QY 781 NSHVRGDDAFKINKDLINATNSNFSURQTKDDFYDGYARNAINSYNTISILGCVTLGG 840  
DB 781 NSHVRGDDAFKINKDLINATNSNFSURQTKDDFYDGYARNAINSYNTISILGCVTLGG 840

QY 841 QNSSSTTGNITTEKAANVTLEANNAPNOQNIIRDVTKLGLSLVNGSLSTGENADIKN 900  
DB 841 QNSSSTTGNITTEKAANVTLEANNAPNOQNIIRDVTKLGLSLVNGSLSTGENADIKN 900

QY 901 LTISEATFKGTRDTLNTNGTNGTAEINTQGVKLGWNTNODGLNITTHAKRNOR 960  
DB 901 LTISEATFKGTRDTLNTNGTNGTAEINTQGVKLGWNTNODGLNITTHAKRNOR 960

QY 961 STIGGDIINKKSLNITDSNNDAAEIQIGGNISOKEGNLTSSDKINITKQITKKGIDGE 1020  
DB 961 STIGGDIINKKSLNITDSNNDAAEIQIGGNISOKEGNLTSSDKINITKQITKKGIDGE 1020

QY 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSAGEAKTV 1080  
DB 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSAGEAKTV 1080

QY 1081 TFNNYKDSKISADGHNVTLSKVKTSNNGGREGNSDNDTGLTITAKNVEVKNKDITSLKT 1140  
DB 1081 TFNNYKDSKISADGHNVTLSKVKTSNNGGREGNSDNDTGLTITAKNVEVKNKDITSLKT 1140

QY 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEA 1200  
DB 1141 VNITASEKVTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEA 1200

QY 1201 KSGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLITEA 1260  
DB 1201 KSGEANTVSATGTIGGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLITEA 1260

QY 1261 GSSITSTKGQVDDLLAQNAGSAGSINAAANVTNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320  
DB 1261 GSSITSTKGQVDDLLAQNAGSAGSINAAANVTNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320

QY 1321 LNGDASGDSDEVNAVNASGSGSVTAATSSSVNTGDLTNGVNLNIIISKOGNRTVRLRGE 1380  
DB 1321 LNGDASGDSDEVNAVNASGSGSVTAATSSSVNTGDLTNGVNLNIIISKOGNRTVRLRGE 1380

QY 1381 IEVKYIQGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRVEPNNTITVNTON 1440  
DB 1381 IEVKYIQGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRVEPNNTITVNTON 1440

QY 1441 EFTTRPSQVILIEGKACFSSGNGARVCTNVADGGQ 1477  
DB 1441 EFTTRPSQVILIEGKACFSSGNGARVCTNVADGGQ 1477

## RESULT 4

US-08-469-880-4  
; Sequence 4, Application US/08469880  
; Patent No. 5876733  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: Of No. 5876733-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,880  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810									
TELEFAX: (703) 415-0813									
INFORMATION FOR SEQ ID NO: 4:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 1477 amino acids									
TYPE: amino acid									
STRANDEDNESS: single									
TOPOLOGY: linear									
US-08-469-880-4									
Query Match									
Best Local Similarity 100.0%; Score 7407; DB 2; Length 1477;									
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTKESKPKARMKVRHLALPKPLSALLSLGVT	60						
DB	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTKESKPKARMKVRHLALPKPLSALLSLGVT	60						
QY	61	SIQSVLASGLQGMVDVHGTATMQVDGNKTIIRNSVDAAIINMKOFNIDQNEVMVQFLOENN	120						
DB	61	SIQSVLASGLQGMVDVHGTATMQVDGNKTIIRNSVDAAIINMKOFNIDQNEVMVQFLOENN	120						
QY	121	NSAVFNRTVTSNOISQKGLDSNGOVFLINPNGIITIGKDAIINTNGFTASTLDSINENIK	180						
DB	121	NSAVFNRTVTSNOISQKGLDSNGOVFLINPNGIITIGKDAIINTNGFTASTLDSINENIK	180						
QY	181	ARNFTFQPKDKALAEIVNHGLITVKGDSVNLIGKVKNEGVI SVNGGSI SLLAGOKIT	240						
DB	181	ARNFTFQPKDKALAEIVNHGLITVKGDSVNLIGKVKNEGVI SVNGGSI SLLAGOKIT	240						
QY	241	ISDINPTTYSIAAPENAEVNLGDI FAKGGNINVRATIRNOGKLSADSVSKDKSGNIV	300						
DB	241	ISDINPTTYSIAAPENAEVNLGDI FAKGGNINVRATIRNOGKLSADSVSKDKSGNIV	300						
QY	301	LSNKEGAEIGVISAOQOAGKGLMITGDKVTLTKTGAVIDLSKGEGETVLLGDERGE	360						
DB	301	LSNKEGAEIGVISAOQOAGKGLMITGDKVTLTKTGAVIDLSKGEGETVLLGDERGE	360						
QY	361	GKNGIOLAKTSLKSTINVSKEKGGFATVWGDIALDGNINAGSGDIAKTGGFVET	420						
DB	361	GKNGIOLAKTSLKSTINVSKEKGGFATVWGDIALDGNINAGSGDIAKTGGFVET	420						
QY	421	SGHDLFIKONAIVDAKEWLLDFDNVSINAEEDPLFNNTGINDPEPTGGEASDPKKNSELK	480						
DB	421	SGHDLFIKONAIVDAKEWLLDFDNVSINAEEDPLFNNTGINDPEPTGGEASDPKKNSELK	480						
QY	481	TTLTNTTISNLYKNAWMTNITASRKLTVNSSINIGSNHLLILHSKGQRCGGVQIDGDITS	540						
DB	481	TTLTNTTISNLYKNAWMTNITASRKLTVNSSINIGSNHLLILHSKGQRCGGVQIDGDITS	540						
QY	541	KGNLTITISGQWVDVHKNTITLDQGLNITAAASVAFEGGNNKARDAANAKIVAOGTVTITG	600						
DB	541	KGNLTITISGQWVDVHKNTITLDQGLNITAAASVAFEGGNNKARDAANAKIVAOGTVTITG	600						
QY	601	EGKDFRANNVSLNGTCKGLNIISVNNLTHNLSGTINISGNIITINOTTRKNTSYWQTSMD	660						
DB	601	EGKDFRANNVSLNGTCKGLNIISVNNLTHNLSGTINISGNIITINOTTRKNTSYWQTSMD	660						
QY	661	SHNNVSALNLETGANFTFIKVISSNSKGLTQYRSSAGVNFNGVNGMNSFNLKEGAKVNF	720						
DB	661	SHNNVSALNLETGANFTFIKVISSNSKGLTQYRSSAGVNFNGVNGMNSFNLKEGAKVNF	720						
QY	721	KLAFENWNTSKPLPFRFLANITATGGSVFFDIYAHNSGRGAELKMSINISNGANFTLL	780						
DB	721	KLAFENWNTSKPLPFRFLANITATGGSVFFDIYAHNSGRGAELKMSINISNGANFTLL	780						
QY	781	NSHVRGDDAFKINKDLTINATNSFSLQTKDFDYDGYARNAINSYNISILGCVNLTGG	840						
DB	781	NSHVRGDDAFKINKDLTINATNSFSLQTKDFDYDGYARNAINSYNISILGCVNLTGG	840						
QY	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIERDVTKLGSLLVNGSLTGTENADIKGN	900						
DB	841	QNSSSITGNITIEKAANVTLEANNAPNOQNIERDVTKLGSLLVNGSLTGTENADIKGN	900						

QY	901	LTISSEATFKGKTRDTLNTITNFTNNGTAEINITOGVVKLGNTVNDGDLNITHAKRNOR	960
DB	901	LTISSEATFKGKTRDTLNTITNFTNNGTAEINITOGVVKLGNTVNDGDLNITHAKRNOR	960
QY	961	SIIGGDIINKKGSNLITDSNNDAAEIOIGGNI SOKEGNLTISSDKINITKQITIKKGIDGE	1020
DB	961	SIIGGDIINKKGSNLITDSNNDAAEIOIGGNI SOKEGNLTISSDKINITKQITIKKGIDGE	1020
QY	1021	DSSSDATSNANLTIKTKELKLTEDLSISGFNKAEITAKDGRDLTIGNSNDGNSGAEAKTV	1080
DB	1021	DSSSDATSNANLTIKTKELKLTEDLSISGFNKAEITAKDGRDLTIGNSNDGNSGAEAKTV	1080
QY	1081	TFNNVKDSKISADGHNVTLASKVKTSSNGGRESNSDNDTGLTITAKNVVKNKDITSLKT	1140
DB	1081	TFNNVKDSKISADGHNVTLASKVKTSSNGGRESNSDNDTGLTITAKNVVKNKDITSLKT	1140
QY	1141	VNITASEKVTITAGSTINATNGKASITTKTGDLSGTISGNTVSVSATVDLTTKSGSKIEA	1200
DB	1141	VNITASEKVTITAGSTINATNGKASITTKTGDLSGTISGNTVSVSATVDLTTKSGSKIEA	1200
QY	1201	KSGEANYTSATGITIGTISGNTVNTANAGDLTVNGCAEINATEGAATLTATGNTLTTEA	1260
DB	1201	KSGEANYTSATGITIGTISGNTVNTANAGDLTVNGCAEINATEGAATLTATGNTLTTEA	1260
QY	1261	GSSITSTKGQVDDLLAONGSIAGSINAANVTNTTGTTLTVAGSDIKATSGTLVINAKDAK	1320
DB	1261	GSSITSTKGQVDDLLAONGSIAGSINAANVTNTTGTTLTVAGSDIKATSGTLVINAKDAK	1320
QY	1321	LNGDASGDSSTEVNNAVNASGSGSVTAATSSSVNITGDLTNTVNGLNIIISKDGRTVRLRGE	1380
DB	1321	LNGDASGDSSTEVNNAVNASGSGSVTAATSSSVNITGDLTNTVNGLNIIISKDGRTVRLRGE	1380
QY	1381	IEVKYIOPGVAESVEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRVEPNNTITVNTON	1440
DB	1381	IEVKYIOPGVAESVEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRVEPNNTITVNTON	1440
QY	1441	EFTTRPSSQVIISGKACFCSSGNGARVCTNNVADDDQGP	1477
DB	1441	EFTTRPSSQVIISGKACFCSSGNGARVCTNNVADDDQGP	1477

RESULT 5  
US-08-728-470-4  
; Sequence 4, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Baronkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728.470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1477 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-4

Query Match 100.0%; Score 7407; DB 2; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKIYRFLSKRLNALVAVSELARGCDHSTKSGSEKPAKMKVRHLALPLSAMLISLQVT 60  
DB 1 MKKIYRFLSKRLNALVAVSELARGCDHSTKSGSEKPAKMKVRHLALPLSAMLISLQVT 60

QY 61 SIPOSVLASGLQMDVHGHTATQMDVGNKTIIRNSVDIINWKFNDQNMWQFLOENN 120  
DB 61 SIPOSVLASGLQMDVHGHTATQMDVGNKTIIRNSVDIINWKFNDQNMWQFLOENN 120

QY 121 NSAVFNRTVNSQISOLKGLDNGSQVFLNPNGITIGKDAIINNGFTASTFLDLSNENIK 180  
DB 121 NSAVFNRTVNSQISOLKGLDNGSQVFLNPNGITIGKDAIINNGFTASTFLDLSNENIK 180

QY 181 ARNFTFQOTKOKALAEIVNHGLIITVGKDSYNLIGKVKNEGIVSVNGGSIISLLAGOKIT 240  
DB 181 ARNFTFQOTKOKALAEIVNHGLIITVGKDSYNLIGKVKNEGIVSVNGGSIISLLAGOKIT 240

QY 241 ISDIINPTIYSIAAPENAEVNLGDIIFAKGGINVRAATIRNOGKLSADSVSKDKSGNIV 300  
DB 241 ISDIINPTIYSIAAPENAEVNLGDIIFAKGGINVRAATIRNOGKLSADSVSKDKSGNIV 300

QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360  
DB 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360

QY 361 KGNGIQLAKKTSLEKGSTINVSQKKEGFAIVMGDIALIDGNINAQSGDIAGTGGFVET 420  
DB 361 KGNGIQLAKKTSLEKGSTINVSQKKEGFAIVMGDIALIDGNINAQSGDIAGTGGFVET 420

QY 421 SGHDLFIKDNAIVDAKEWLLDFDNVSIADPLFNNTGINDPPTGTGEASDPKKNSELK 480  
DB 421 SGHDLFIKDNAIVDAKEWLLDFDNVSIADPLFNNTGINDPPTGTGEASDPKKNSELK 480

QY 481 TLTNTTISYLNKAWMTNTASRKLTVNSNIGNSHLILHSKQORGQVQIDGDIITS 540  
DB 481 TLTNTTISYLNKAWMTNTASRKLTVNSNIGNSHLILHSKQORGQVQIDGDIITS 540

QY 541 KGNLTIYSGGWVDVHKNTIDQGLFNITAAVAFEGGNKARDAANAKIQAQGTITNG 600  
DB 541 KGNLTIYSGGWVDVHKNTIDQGLFNITAAVAFEGGNKARDAANAKIQAQGTITNG 600

QY 601 EGKDFRANVNSLNGTKGLNISSVNNLTHNLSTGTTINISGNITINQTRKNTSYWQTSHD 660  
DB 601 EGKDFRANVNSLNGTKGLNISSVNNLTHNLSTGTTINISGNITINQTRKNTSYWQTSHD 660

QY 661 SHNVNLSALNLETGANFTFKIYSSNSKGLTQYRSSAGVFNFGVNGNNSFNLKEGAKVNF 720  
DB 661 SHNVNLSALNLETGANFTFKIYSSNSKGLTQYRSSAGVFNFGVNGNNSFNLKEGAKVNF 720

QY 721 KLPKNENMNTSKPLPIRFANITATGGGVFFDIYANHSGRGAELKMSSEINISNCANFTL 780  
DB 721 KLPKNENMNTSKPLPIRFANITATGGGVFFDIYANHSGRGAELKMSSEINISNCANFTL 780

DB 721 KLPKNENMNTSKPLPIRFANITATGGGVFFDIYANHSGRGAELKMSSEINISNCANFTL 780  
QY 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNNVTLGG 840  
DB 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNNVTLGG 840  
QY 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIROVRVTKLGLSVLLVNGSLTGENADIKGN 900  
DB 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIROVRVTKLGLSVLLVNGSLTGENADIKGN 900  
QY 901 LTISESATFKGTRDTLNIITGNFTNNGTAEINITQGVVVKLGNVNDGDLNITTHAKNQR 960  
DB 901 LTISESATFKGTRDTLNIITGNFTNNGTAEINITQGVVVKLGNVNDGDLNITTHAKNQR 960  
QY 961 SIIGGDIINKGSLNITDSNNDAEIQIIGGNIQKEGNLTISSDKINITKQITIKKGIDGE 1020  
DB 961 SIIGGDIINKGSLNITDSNNDAEIQIIGGNIQKEGNLTISSDKINITKQITIKKGIDGE 1020  
QY 1021 DSSSDATSNANLTITKTRELKLTEDLSISGFNKAEITAKDGRDLTIGNSDGNSGAEAKTV 1080  
DB 1021 DSSSDATSNANLTITKTRELKLTEDLSISGFNKAEITAKDGRDLTIGNSDGNSGAEAKTV 1080  
QY 1081 TFNNVDSKISADGHNVTLSKVKTSNNGSGRESNSDNDTGLTITAKNVENKDIITSLKT 1140  
DB 1081 TFNNVDSKISADGHNVTLSKVKTSNNGSGRESNSDNDTGLTITAKNVENKDIITSLKT 1140  
QY 1141 VNITASEKVTITAGSTINATNGKASITTKTGDIISGTISGNTVSVSATVDLTTKSGSKIEA 1200  
DB 1141 VNITASEKVTITAGSTINATNGKASITTKTGDIISGTISGNTVSVSATVDLTTKSGSKIEA 1200  
QY 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260  
DB 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260  
QY 1261 GSSITSTKGOVDLLAONGSTAGSINAAVNTLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320  
DB 1261 GSSITSTKGOVDLLAONGSTAGSINAAVNTLNTTGTTLTVAGSDIKATSGTLVINAKDAK 1320  
QY 1321 LNGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNIISKDGRNTVRLRGKE 1380  
DB 1321 LNGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNIISKDGRNTVRLRGKE 1380  
QY 1381 IEVKYIQQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVNTQN 1440  
DB 1381 IEVKYIQQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRFPENNTITVNTQN 1440  
QY 1441 EFTTRPSSQVILISEGKACFSNGNGARVCTNVAADDGQP 1477  
DB 1441 EFTTRPSSQVILISEGKACFSNGNGARVCTNVAADDGQP 1477

## RESULT 6

US-08-617-697-4  
; Sequence 4, Application US/08617697  
; Patent No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; City: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

Query Match 100.0%; Score 7407; DB 2; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLGLV 60  
DB 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLGLV 60  
QY 61 SIPOSVLASGLQGVVHGTATMGVGNKTIIRNSVDALINWKFNDQEMVQFQLENN 120  
DB 61 SIPOSVLASGLQGVVHGTATMGVGNKTIIRNSVDALINWKFNDQEMVQFQLENN 120  
QY 121 NSAVFNRYTSNQISOLKGLDSNGQVFLINPNGITIGKDALINNGFTASTLDSNENIK 180  
DB 121 NSAVFNRYTSNQISOLKGLDSNGQVFLINPNGITIGKDALINNGFTASTLDSNENIK 180  
QY 181 ARNFTFQTKDKALAEIIVNHGLITVKGDSVNLIGGKVKNEGVSIVNGGSISSLAGOKIT 240  
DB 181 ARNFTFQTKDKALAEIIVNHGLITVKGDSVNLIGGKVKNEGVSIVNGGSISSLAGOKIT 240  
QY 241 ISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSYSKDKSGNIV 300  
DB 241 ISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSYSKDKSGNIV 300  
QY 301 LSARGEAFIGVISAQNOQAKGGLMITGDKVTLTKCAVTDLSGKEGGTYLGGDERGE 360  
DB 301 LSARGEAFIGVISAQNOQAKGGLMITGDKVTLTKCAVTDLSGKEGGTYLGGDERGE 360  
QY 361 GRNGQLAKKTSLEKGSITINVSKEKGFIVWGDIALIDGNINAQSGDIAKTGGFVET 420  
DB 361 GRNGQLAKKTSLEKGSITINVSKEKGFIVWGDIALIDGNINAQSGDIAKTGGFVET 420  
QY 421 SGHDLFIKDNAIVDAKEWLLDFNDVSNINAEPLFNNTGINDPEFTGTGEASDPKKNSELK 480  
DB 421 SGHDLFIKDNAIVDAKEWLLDFNDVSNINAEPLFNNTGINDPEFTGTGEASDPKKNSELK 480  
QY 481 TLTNTNTISNYLKNWMTNMTASRLKTLVNSSINIGNSHLILHSKQGGGVQIDGDITS 540  
DB 481 TLTNTNTISNYLKNWMTNMTASRLKTLVNSSINIGNSHLILHSKQGGGVQIDGDITS 540  
QY 541 KGNLTIYSGGVVHVHKNITLDQGFNLITAAVAPEGGNKNKARDAANAKIVAQGTVTITG 600  
DB 541 KGNLTIYSGGVVHVHKNITLDQGFNLITAAVAPEGGNKNKARDAANAKIVAQGTVTITG 600  
QY 601 EGKDFRANNVSLNCTCKGLNLISSVNLTHNLSGTINISGNITINOTTRKNTSYQWTSID 660  
DB 601 EGKDFRANNVSLNCTCKGLNLISSVNLTHNLSGTINISGNITINOTTRKNTSYQWTSID 660

US-08-617-697-4

Query Match 100.0%; Score 7407; DB 2; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLGLV 60  
DB 1 MNKYRLKFKRLNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSAMLGLV 60  
QY 61 SIPOSVLASGLQGVVHGTATMGVGNKTIIRNSVDALINWKFNDQEMVQFQLENN 120  
DB 61 SIPOSVLASGLQGVVHGTATMGVGNKTIIRNSVDALINWKFNDQEMVQFQLENN 120  
QY 121 NSAVFNRYTSNQISOLKGLDSNGQVFLINPNGITIGKDALINNGFTASTLDSNENIK 180  
DB 121 NSAVFNRYTSNQISOLKGLDSNGQVFLINPNGITIGKDALINNGFTASTLDSNENIK 180  
QY 181 ARNFTFQTKDKALAEIIVNHGLITVKGDSVNLIGGKVKNEGVSIVNGGSISSLAGOKIT 240  
DB 181 ARNFTFQTKDKALAEIIVNHGLITVKGDSVNLIGGKVKNEGVSIVNGGSISSLAGOKIT 240  
QY 241 ISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSYSKDKSGNIV 300  
DB 241 ISDIINPTITYSIAAPENEAVALNGLDIFAKGGINVRAATIRNOGKLSADSYSKDKSGNIV 300  
QY 301 LSARGEAFIGVISAQNOQAKGGLMITGDKVTLTKCAVTDLSGKEGGTYLGGDERGE 360  
DB 301 LSARGEAFIGVISAQNOQAKGGLMITGDKVTLTKCAVTDLSGKEGGTYLGGDERGE 360  
QY 361 GRNGQLAKKTSLEKGSITINVSKEKGFIVWGDIALIDGNINAQSGDIAKTGGFVET 420  
DB 361 GRNGQLAKKTSLEKGSITINVSKEKGFIVWGDIALIDGNINAQSGDIAKTGGFVET 420  
QY 421 SGHDLFIKDNAIVDAKEWLLDFNDVSNINAEPLFNNTGINDPEFTGTGEASDPKKNSELK 480  
DB 421 SGHDLFIKDNAIVDAKEWLLDFNDVSNINAEPLFNNTGINDPEFTGTGEASDPKKNSELK 480  
QY 481 TLTNTNTISNYLKNWMTNMTASRLKTLVNSSINIGNSHLILHSKQGGGVQIDGDITS 540  
DB 481 TLTNTNTISNYLKNWMTNMTASRLKTLVNSSINIGNSHLILHSKQGGGVQIDGDITS 540  
QY 541 KGNLTIYSGGVVHVHKNITLDQGFNLITAAVAPEGGNKNKARDAANAKIVAQGTVTITG 600  
DB 541 KGNLTIYSGGVVHVHKNITLDQGFNLITAAVAPEGGNKNKARDAANAKIVAQGTVTITG 600  
QY 601 EGKDFRANNVSLNCTCKGLNLISSVNLTHNLSGTINISGNITINOTTRKNTSYQWTSID 660  
DB 601 EGKDFRANNVSLNCTCKGLNLISSVNLTHNLSGTINISGNITINOTTRKNTSYQWTSID 660

US-08-617-697-4

QY 661 SHWNVSALENTGANTFTPIKYISSNKGTLTQYRSSAGVNFNGVNMSPFLKGEKAVNF 720  
DB 661 SHWNVSALENTGANTFTPIKYISSNKGTLTQYRSSAGVNFNGVNMSPFLKGEKAVNF 720  
QY 721 KLKPNENMTSKPLPIRELANITATGGGVFFDIYANHSGRGAELKXSEINISNGANFTL 780  
DB 721 KLKPNENMTSKPLPIRELANITATGGGVFFDIYANHSGRGAELKXSEINISNGANFTL 780  
QY 781 NSHVGGDDAFKINKDLTITNATNSFSLQTKDDFFDGYARNAINSTYNSILGNGVTLGG 840  
DB 781 NSHVGGDDAFKINKDLTITNATNSFSLQTKDDFFDGYARNAINSTYNSILGNGVTLGG 840  
QY 841 QNSSSSITGNTPIEKAANVTLEANNAPNOONIRDRVYIKGLSLLVNGSLTGERADIKGN 900  
DB 841 QNSSSSITGNTPIEKAANVTLEANNAPNOONIRDRVYIKGLSLLVNGSLTGERADIKGN 900  
QY 901 LTISESATFKGKTRDTLNTGNTNGTAEINITOGVVKLGNVTNDGDLNITTHAKNRQR 960  
DB 901 LTISESATFKGKTRDTLNTGNTNGTAEINITOGVVKLGNVTNDGDLNITTHAKNRQR 960  
QY 961 SIIGDDIINKGSLNITDSNNDAEIOIGNISOKEGNLTISSDKINITKQITIKKIDGGE 1020  
DB 961 SIIGDDIINKGSLNITDSNNDAEIOIGNISOKEGNLTISSDKINITKQITIKKIDGGE 1020  
QY 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDCRDLTIGNSNDGSGAEAKTV 1080  
DB 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDCRDLTIGNSNDGSGAEAKTV 1080  
QY 1081 TPNNVKDSKISADGHNVTLSNKTSSNGRESNDSNDTGLTITAKNVEVKNKDTISLKT 1140  
DB 1081 TPNNVKDSKISADGHNVTLSNKTSSNGRESNDSNDTGLTITAKNVEVKNKDTISLKT 1140  
QY 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTSISGTSVSATVDLTITKSGSKIEA 1200  
DB 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISGTSISGTSVSATVDLTITKSGSKIEA 1200  
QY 1201 KSGEANTVATGTTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260  
DB 1201 KSGEANTVATGTTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260  
QY 1261 GSSITSTKGVDDLLAONGSITAGSINAAVNTLTGTLTVAGSDIKATSGTLVINAKDAK 1320  
DB 1261 GSSITSTKGVDDLLAONGSITAGSINAAVNTLTGTLTVAGSDIKATSGTLVINAKDAK 1320  
QY 1321 LNAGDASGDSPEVNAVNASGSGSVTAATSSSVNTITGDLTVNGLNLISSKOGRTVRLRGE 1380  
DB 1321 LNAGDASGDSPEVNAVNASGSGSVTAATSSSVNTITGDLTVNGLNLISSKOGRTVRLRGE 1380  
QY 1381 IEVYIOPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRPEVPEPNTITVNTON 1440  
DB 1381 IEVYIOPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLGVSAVRPEVPEPNTITVNTON 1440  
QY 1441 EFTTRPSSQVLIISBKGACFSGNGARVCTNVADDGQP 1477  
DB 1441 EFTTRPSSQVLIISBKGACFSGNGARVCTNVADDGQP 1477

## RESULT 7

US-08-719-641-4  
; Sequence 4, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.



```

; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-4

```

```

Query Match      100.0%; Score 7407; DB 4; Length 1477;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKIYRLKSKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALPLSAMLISLQVT 60
DB 1 MNKIYRLKSKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALPLSAMLISLQVT 60

QY 61 SIPOSVLASGLQGMVDVHGVTATMDVGNKTIIRNSVDIAIINWKQFNIDQNEWQFLOENN 120
DB 61 SIPOSVLASGLQGMVDVHGVTATMDVGNKTIIRNSVDIAIINWKQFNIDQNEWQFLOENN 120

QY 121 NSAVFNRYTSNQISQLKGLTDSNGOVFLINPNGITIGKDAIINTNGFTASTLDIISNENIK 180
DB 121 NSAVFNRYTSNQISQLKGLTDSNGOVFLINPNGITIGKDAIINTNGFTASTLDIISNENIK 180

QY 181 ARNFTFEQTKKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGGSIISLLAGOKIT 240
DB 181 ARNFTFEQTKKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGGSIISLLAGOKIT 240

QY 241 ISDIINPTITYSIAAPENEAIVNGLDIFAKGGINIVRAATIRNOGKLSADSVSKDKSGNIV 300
DB 241 ISDIINPTITYSIAAPENEAIVNGLDIFAKGGINIVRAATIRNOGKLSADSVSKDKSGNIV 300

QY 301 LSAKEGEAEIGGVISAQOQAGKGLMITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360
DB 301 LSAKEGEAEIGGVISAQOQAGKGLMITGDKVTLTKTGAVIDLSGKEGGTYLGGDERGE 360

QY 361 CKNGIOLAKKTSLEKGSTINVSKEGKGFAIVWGDIALIDGNINAOQSGDIAGTGGFVET 420
DB 361 CKNGIOLAKKTSLEKGSTINVSKEGKGFAIVWGDIALIDGNINAOQSGDIAGTGGFVET 420

QY 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGINDERPTGTGEASDPKKNSELK 480
DB 421 SGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGINDERPTGTGEASDPKKNSELK 480

QY 481 TLTNTTISNVLKNAWTMTNITASRKLTVNNSINIGNSHLILHSKQGGGVQIDGDITS 540
DB 481 TLTNTTISNVLKNAWTMTNITASRKLTVNNSINIGNSHLILHSKQGGGVQIDGDITS 540

```

```

QY 541 KCGNLTIYSGGWVDVHKNTLDQGFNLITAAVAFEGGNKARDAANAKIVAQGVTTITG 600
DB 541 KCGNLTIYSGGWVDVHKNTLDQGFNLITAAVAFEGGNKARDAANAKIVAQGVTTITG 600

QY 601 EGDFRANNVSLNGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTKNTSYQWTSID 660
DB 601 EGDFRANNVSLNGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTKNTSYQWTSID 660

QY 661 SHWNVSALENGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTKNTSYQWTSID 720
DB 661 SHWNVSALENGTGKGLNIISVNNLTHNLSGTINISGNITINQTTTKNTSYQWTSID 720

QY 721 KLKPNENMTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMEISINSGANFTL 780
DB 721 KLKPNENMTSKPLPIRFLANITATGGGVFFDIYANHSGRGAELKMEISINSGANFTL 780

QY 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGVARNAINSTYISILGGNVTILGG 840
DB 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGVARNAINSTYISILGGNVTILGG 840

QY 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDRIKGLSLLVNGSLTGENADIKGN 900
DB 841 QNSSSITGNITIEKAANVTLEANNAPNOQNIIRDRIKGLSLLVNGSLTGENADIKGN 900

QY 901 LTISESATFKGTRDTLNTGNFTNNGTAEINITQGVVVKLGNTVDGDLNITTHAKRNOR 960
DB 901 LTISESATFKGTRDTLNTGNFTNNGTAEINITQGVVVKLGNTVDGDLNITTHAKRNOR 960

QY 961 SIIGDIIINKKGSNITDSNNDAEIQIGGNIQKEGNLITSSDKINITKQITIKKIDGE 1020
DB 961 SIIGDIIINKKGSNITDSNNDAEIQIGGNIQKEGNLITSSDKINITKQITIKKIDGE 1020

QY 1021 DSSSDATSNANLTIKTRELKLTEDLSISGFNKAETITAKDGRDLTIGNSDGNSGAEAKTV 1080
DB 1021 DSSSDATSNANLTIKTRELKLTEDLSISGFNKAETITAKDGRDLTIGNSDGNSGAEAKTV 1080

QY 1081 TFNNVKOSKISADGHNVTLNSKVKTSNNGRESNNDGTGLTTAKNVEVNDITSLKT 1140
DB 1081 TFNNVKOSKISADGHNVTLNSKVKTSNNGRESNNDGTGLTTAKNVEVNDITSLKT 1140

QY 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTLISGNTVSVSATVDLTTSKSGKIEA 1200
DB 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTLISGNTVSVSATVDLTTSKSGKIEA 1200

QY 1201 KSGEANVTSAITGIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260
DB 1201 KSGEANVTSAITGIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260

QY 1261 GSSITSTKGOVDLLAQNAGSTAGSINAAVNTLNTTGLTTVAGSDIKATSGTLVINAKDAK 1320
DB 1261 GSSITSTKGOVDLLAQNAGSTAGSINAAVNTLNTTGLTTVAGSDIKATSGTLVINAKDAK 1320

QY 1321 LNGDASGDSPEVNAVNASGSGSVTAATSSVNTTGDLTNVNGLNIISKDGRNTRVRURKE 1380
DB 1321 LNGDASGDSPEVNAVNASGSGSVTAATSSVNTTGDLTNVNGLNIISKDGRNTRVRURKE 1380

QY 1381 IEVKYIQPGVASVEEYIEAKRVLEKVKDLSDEERETLAKLGVSAVRFVPEPNTITVNTON 1440
DB 1381 IEVKYIQPGVASVEEYIEAKRVLEKVKDLSDEERETLAKLGVSAVRFVPEPNTITVNTON 1440

QY 1441 EFTTRPSSQVILISEGKACFSNGNARVCNTVADDDQOP 1477
DB 1441 EFTTRPSSQVILISEGKACFSNGNARVCNTVADDDQOP 1477

```

```

RESULT 8
US-09-206-942-71
; Sequence 71, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena M.
; APPLICANT: Yang, Yan-Ping

```

```

; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: Molecular Weight Proteins
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 71
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-206-942-71

Query Match      98.0%; Score 7261; DB 4; Length 1477;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1454; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

Qy 1 MNKIYRLKFSKRLNALVANSELARGCDHSTKSEKSPARKVPHLALKPLSAMLISLGV 60
Db 1 MNKIYRLKFSKRLNALVANSELARGCDHSTKSEKFRVYTFRCNHLALKPLSAMLISLGV 60
Qy 61 SIPOSVLASLGQMDVVGHTATVOVGNKTIIRNSVDAILNKKOFNIDQENWVQFOENN 120
Db 61 SIPOSVLASLGQMDVVGHTATVOVGNKTIIRNSVDAILNKKOFNIDQENWVQFOENN 120
Qy 121 NSAVFNRTVNSOISQLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
Db 121 NSAVFNRTVNSOISQLKGLDSNGQVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
Qy 181 ARNFTFQTKDKALAEIVNHLITVKGDSVNLIGGKVRNEGVIYVNGGSIISLAGOKIT 240
Db 181 ARNFTFQTKDKALAEIVNHLITVKGDSVNLIGGKVRNEGVIYVNGGSIISLAGOKIT 240
Qy 241 ISDIINPTIYSAAPENEAVALNLDIFAKGGINVRAATIRNOGKLSADSVSKDKSGNIV 300
Db 241 ISDIINPTIYSAAPENEAVALNLDIFAKGGINVRAATIRNOGKLSADSVSKDKSGNIV 300
Qy 301 LSAKEGEAEIGGVISAQNOAKGKLMITGDKVTLTKTGAVIDLSGKEGTYLGGDRGE 360
Db 301 LSAKEGEAEIGGVISAQNOAKGKLMITGDKVTLTKTGAVIDLSGKEGTYLGGDRGE 360
Qy 361 KNGIOLAKTSLKSGSTINVSKEKGFPATVWGDIADIGNINAQSGDIAGTGGFVET 420
Db 361 KNGIOLAKTSLKSGSTINVSKEKGFPATVWGDIADIGNINAQSGDIAGTGGFVET 420
Qy 421 SGHDLFTKDAIVDAKEMLLDFDWNSTNAEDPLENNNTGINDPEPTGTGEASDPKKNSELK 480
Db 421 SGHDLFTKDAIVDAKEMLLDFDWNSTNAEDPLENNNTGINDPEPTGTGEASDPKKNSELK 480
Qy 481 TTLNTTISNLYLKNAWTMNITASRLKLVNNSINIGSNHSLHLSKRGQGGVQIDGITS 540
Db 481 TTLNTTISNLYLKNAWTMNITASRLKLVNNSINIGSNHSLHLSKRGQGGVQIDGITS 540
Qy 541 KGGNLTYSGGWVDVHKNTIDQGFNLITAASVAFEGGNKARDAANAIAQGVTVITG 600
Db 541 KGGNLTYSGGWVDVHKNTIDQGFNLITAASVAFEGGNKARDAANAIAQGVTVITG 600
Qy 601 EKGDFRANNSVLTGKGLNISSVNNLTNLSGTINISGNITINOTTRKNTSYWQTSHD 660
Db 601 EKGDFRANNSVLTGKGLNISSVNNLTNLSGTINISGNITINOTTRKNTSYWQTSHD 660
Qy 661 SHWVNSALNLETGANFTFKIYSSNSKGLTTOYRSSAGVNFNGVNGNMFNLKGAQVNF 720
Db 661 SHWVNSALNLETGANFTFKIYSSNSKGLTTOYRSSAGVNFNGVNGNMFNLKGAQVNF 720
Qy 721 KUKPENNTSKPLDIRFLANITATGGGVSFFDIYANHSRGAELKUKSEINISGANFTL 780
Db 721 KUKPENNTSKPLDIRFLANITATGGGVSFFDIYANHSRGAELKUKSEINISGANFTL 780
Qy 781 NSHVRGDDAFKINKDLTINATNSNFSRLQTKDDFDYDGYARNAINSTYNISILGGNVLGG 840

```

```

Db 781 NSHVRGDDAFKINKDLTINATNSNFSRLQTKDDFDYDGYARNAINSTYNISILGGNVLGG 840
Qy 841 QNSSSSITGNITIEKAANVTLEANNAPNOQNIROVRVYKLSGLLVNLSLUTGENADIKGN 900
Db 841 QNSSSSITGNITIEKAANVTLEANNAPNOQNIROVRVYKLSGLLVNLSLUTGENADIKGN 900
Qy 901 LTISSEATFKGTRDLTNITGNFTNNGPAEINITQGVVKLGNVNDGDLNITTHIAKNQR 960
Db 901 LTISSEATFKGTRDLTNITGNFTNNGPAEINITQGVVKLGNVNDGDLNITTHIAKNQR 960
Qy 961 SIIGDIIINKKGLNITDSNDAEIQIGNISQKEGNTIISDKINTKQITIKKIDGE 1020
Db 961 SIIGDIIINKKGLNITDSNDAEIQIGNISQKEGNTIISDKINTKQITIKKIDGE 1020
Qy 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGNSGAETV 1080
Db 1021 DSSSDATSNANLTIKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGNSGAETV 1080
Qy 1081 TENNVKDSKISADGHNVTLNSKVTSSNGGRESNDNDTGLTITAKNVEVKNKOITSLKT 1140
Db 1081 TENNVKDSKISADGHNVTLNSKVTSSNGGRESNDNDTGLTITAKNVEVKNKOITSLKT 1140
Qy 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISCTISGNTVSVSATVDLTITKSGSKEA 1200
Db 1141 VNITASEKVTITAGSTINATNGKASITTKTGDISCTISGNTVSVSATVDLTITKSGSKEA 1200
Qy 1201 KSGEANVTSATGTTIGTISGNTVNVATANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
Db 1201 KSGEANVTSATGTTIGTISGNTVNVATANAGDLTVGNGAEINATEGAATLTATGNTLTTEA 1260
Qy 1261 GSSITSTKQVOLLQAQNGSIAGSINAANVTLTNTGTLTTVAGSDIKATSGILVINAKDAK 1320
Db 1261 GSSITSTKQVOLLQAQNGSIAGSINAANVTLTNTGTLTTVAGSDIKATSGILVINAKDAK 1320
Qy 1321 LNGDASGDSSTEVNAVNASGSGSVTAATSSVNTIGDLNTVNGLINIISKDGRNVTBLRKE 1380
Db 1321 LNGDASGDSSTEVNAVNASGSGSVTAATSSVNTIGDLNTVNGLINIISKDGRNVTBLRKE 1380
Qy 1381 IEVKYIQPGVASVEEVIEAKRVLEKVKDISDEERETLAKLGVSAVRFPENPTITVNTQN 1440
Db 1381 IEVKYIQPGVASVEEVIEAKRVLEKVKDISDEERETLAKLGVSAVRFPENPTITVNTQN 1440
Qy 1441 EFTTRPSSQVITISEGKACFSSGNGARVCTNVADDDGQP 1477
Db 1441 EFTTRPSSQVITISEGKACFSSGNGARVCTNVADDDGQP 1477

RESULT 9
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996

```

CLASSIFICATION: 424  
; PRIORITY APPLICATION DATA: US 08/302,832  
; FILING DATE: 05-OCT-1994  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-557  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1600 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-617-697-10

Query Match 71.3%; Score 5282; DB 2; Length 1600;  
Best Local Similarity 68.3%; Pred. No. 1.1e-292;  
Matches 1099; Conservative 136; Mismatches 231; Indels 142; Gaps 16;

QY 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALPLSALLSLGVT 60  
DB 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKSGSEKPKRMKVRHLALPLSALLSLGMA 60

QY 61 SIPOSVLASGLQGMVDVHGTATQMDVGNKTIIRNSVDIAIINNKQFNIDQNMVQFLOENN 120  
DB 61 SIPOSVLASGLQGMVDVHGTATQMDVGNKTIIRNSVDIAIINNKQFNIDQNMVQFLOESS 120

QY 121 NSAVENRVTNQISQKGLDSNGOVFLINPNGITIGCKDAIINNGTFASTLDISNENIK 180  
DB 121 NSAVENRVTNQISQKGLDSNGOVFLINPNGITIGCKDAIINNGTFASTLDISNENIK 180

QY 181 ARNFFTEQTKDALKAEIYVHGLITVKGDSVNLGGKVKNEGIVSVNGGSIISLLAGOKIT 240  
DB 181 ARNFFTEQTKDALKAEIYVHGLITVKGDSVNLGGKVKNEGIVSVNGGSIISLLAGOKIT 240

QY 241 ISDIINPTITYSIAPENEAIVNLGDIKAGGNINVRATIRNQQKLSADSVSKDKSGNIV 300  
DB 241 ISDIINPTITYSIAPENEAIVNLGDIKAGGNINVRATIRNQQKLSADSVSKDKSGNIV 300

QY 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTAVIDLSGKEGGTYLGGDERGE 360  
DB 301 LSAKEGEAEIGGVISAQNOQAKGKLMITGDKVTLTKTAVIDLSGKEGGTYLGGDERGE 360

QY 361 GKNGIOLAKTSLBKGSTINVSKEGKGFAIVMGDIALIDGNINAQGS-DIAKTGGFVET 420  
DB 361 GKNGIOLAKTSLBKGSTINVSKEGKGFAIVMGDIALIDGNINAQGS-DIAKTGGFVET 419

QY 421 SGHDLFTKDAIVDAKELWLDLDFONVSNIAEDPLFNNTGIDNDEPTGTGEASDPKNSLK 480  
DB 420 SGHDLFTKDAIVDAKELWLDLDFONVSNIAEDPLFNNTGIDNDEPTGTGEASDPKNSLK 479

QY 481 TTLNTTISYLNKNAWMTNITASKRLTVSSINIGSNLSHLHSKQGGGVQIDGDTIS 540  
DB 480 TTLNTTISYLNKNAWMTNITASKRLTVSSINIGSNLSHLHSKQGGGVQIDGDTIS 535

QY 541 -KGNLIIYSGWVDVHKNTLTQGGFLNITAA-SVAFEGGNKARDAANAIVAOQTIV 598  
DB 536 NENGLTIKAGSWVDVHKNTLTQGGFLNITAA-SVAFEGGNKARDAANAIVAOQTIV 595

QY 599 TGECKDFRANNSLNGTGKGLNIISSVNNLTNLSCTINISGNTIINQTRKNTSWQTS 658  
DB 596 NKDDQFRFNNSVSLGTGKGLKFIANNQNTFHFDGEINISGIVTINQTKKDVKNWAS 655

QY 659 HDGHNWVSNLNETGANFTFIKYISSNKG--LTTQYRSSAGVNFNGVNGNMSFNLKEGA 716  
DB 656 KDSYWNVSSLTNTVQKFTFIKEVDGSGNQDLRRSSRRSFAGVHFNIGGKTNFNIGANA 715

QY 717 KVNFKLPNENMNTSKPLRFLANITATGG--GSVFDDIYANHSGRAEUKMSEINISN 774  
DB 716 KALFKLPNAAOTPKKELPITFNANITATGNSDSSVDFDIHANILTSRAAGINMDSINTG 775

QY 775 GANFTLNSHVRGDDAFKINKDLTINATNSFSLRQTKDDFDYGVARNATSYNISILGG 834  
DB 776 GLDFSITSHRNSNAFEIKKDLTINATGNSFSLRQTKDDFDYGVARNATSYNISILGG 835

QY 835 NYTLGGNSSSSITGNITIEKAAVTLLEANNAPQOQNIIRDVIKLGLSVNGSLSLTGEN 894  
DB 836 NYTLGGNSSSSITGNITINKANVTLOADTSNSNTGLKRTLTPLGNSVEGNSLSTGAN 895

QY 895 ADIKGNLTISEATEFKGKTRDTLNTGNFTNNGTAEINITQGVVYKLVNVTGDLNITTH 954  
DB 896 ANIVGNLSIAEDSTFKGEASDNLTITFTNNGTANIN-KGVVYKLVNVTGDLNITTH 954

QY 955 AKRNQRIIGDIIINKGSLNITSDNDAEIQIGGNISQKEGNLTSSDKINITKQITIK 1014  
DB 955 ASGTQKTIINGNITNEKGDNLNINIKADAEIQIGGNISQKEGNLTSSDKINITKQITIK 1014

QY 1015 KGIDGEDSSSDATSNANLTITKELKLTEDLSISGFNKAIEITAKDGRDLTIGNSDNGSG 1074  
DB 1015 AGVEGGRSDSEAEANLTIQTKELKLTEDLSISGFNKAIEITAKDGRDLTIGNSDNGSG 1072

QY 1075 AEAKTVTFNNVKDSKISADGHNVTLSKVKVTSNNGSGRESNDSNDTGLTITAKNVEVND 1134  
DB 1073 ADARKVTFDVKDKSISTDGHVNTLSEVKT--SNSSNAGNDSNGLTISAKDVTVNN 1130

QY 1135 ITSLSKTVNI-----TASEKVTIT 1152  
DB 1131 VTSKHTIINSAAGNVTTKEGTINATGTSVEVTAQNGTITKGNITSQNVTVTATENLVTT 1190

QY 1153 AGSTINATNGKASITTKTGDISG-----TISGNTVSVS-----ATVD---L 1190  
DB 1191 ENAVINATSGTVNISVTKGDIKGIESTSGNVNITASGNTLKVSNITGQDVTVTADAGAL 1250

QY 1191 TTKSGSKIEAKSEANVTSATG-----TISGNTVSVS-----ATVD---L 1212  
DB 1251 TTTAGTISATTTGNANITTKTGDISGKVSSESVTLVATGATLAVGNISGNTVTITADS 1310

QY 1213 -----TIG-----GTISGNTVNTANAGDLTVGCAEINATEGAAT 1248  
DB 1311 GKLTISVGTINGTNSVTTSSQSGDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAAT 1370

QY 1249 LRATGNLTTEAGSSITSTKGQVDDLLAQNAGSIAGSINAAVNTLNTGLTITVAGSDIKAT 1308  
DB 1371 LTAESKLTQTGSSITSSNGQTTLTAKDSSIAGNINAANVTLNTGLTITVAGSDIKAT 1430

QY 1309 SGFLVINAKDAKLNGDASGDSFEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNITISK 1368  
DB 1431 SGFLVINAKDAKLNGDASGDSFEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNITISK 1490

QY 1369 DGRNTVRLRGKEIEVKYIQPGVASVEEVEIAKRVLEKVKDLSEBERETLAKLGVSAVRPV 1428  
DB 1491 NGRNTVRLRGKEIEVKYIQPGVASVEEVEIAKRVLEKVKDLSEBERETLAKLGVSAVRPV 1550

QY 1429 EPNNITVNTQNEFTTRPSSQVLIISSEKACFSNGARVCTNVADGG 1476  
DB 1551 EPNNITVNTQNEFTTRPSSQVLIISSEKACFSNGARVCTNVADGG 1598

## RESULT 10

US-09-206-942-73  
; Sequence 73, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb

Tue Mar 25 08:37:45 2003

```
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 73
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-73

Query Match      69.7%; Score 5164; DB 4; Length 1036;
Best Local Similarity 99.5%; Pred. No. 3.4e-286;
Matches 1030; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 443 DNVSINAEDPLFNNTGINDPEFTGTGEASDPKKNSELKTTTLTNTTISNYLKNAMTMTNITA 502
DB 2 DQVTEAEDPLRNNTGINDPEFTGTGEASDPKKNSELKTTTLTNTTISNYLKNAMTMTNITA 61
QY 503 SRKLTWNSSINIGSNHLLHSHKGGGQVQIDGDTISKGGNLTIIYSGGWDVHVKNTILD 562
DB 62 SRKLTWNSSINIGSNHLLHSHKGGGQVQIDGDTISKGGNLTIIYSGGWDVHVKNTILD 121
QY 563 QGFLNTAASVAFEGGNKARDAANAKIVAQGTVTITCEGKDFRANNVSLNCTGKGLNII 622
DB 122 QGFLNTAASVAFEGGNKARDAANAKIVAQGTVTITCEGKDFRANNVSLNCTGKGLNII 181
QY 623 SSVNNLTHNLSGTINISGNITNTOTTRKNTSYWQTSWSDHSHWNSALNLETGANFTFIKYI 682
DB 182 SSVNNLTHNLSGTINISGNITNTOTTRKNTSYWQTSWSDHSHWNSALNLETGANFTFIKYI 241
QY 683 SSNSKGLTTQYRSAGVNFNGVNGNSFNLKEGAKVFNKLPKNENMNTSKPLPIRFLANI 742
DB 242 SSNSKGLTTQYRSAGVNFNGVNGNSFNLKEGAKVFNKLPKNENMNTSKPLPIRFLANI 301
QY 743 TATGGGSVFDDIYANHSRGAELKMEIINISGANFTLNSHVRGDDAFKINKDLTINATN 802
DB 302 TATGGGSVFDDIYANHSRGAELKMEIINISGANFTLNSHVRGDDAFKINKDLTINATN 361
QY 803 SNFLRQTKDDFYDYGARNAINTSYNISILGNGVNTILGGNSSSSITGNITIEKAANVTLE 862
DB 362 SNFLRQTKDDFYDYGARNAINTSYNISILGNGVNTILGGNSSSSITGNITIEKAANVTLE 421
QY 863 ANNAPNOQTRDRVILKGLSVNGSLTGENADIKGNLTISESATFKGTRDTLTNITGN 922
DB 422 ANNAPNOQTRDRVILKGLSVNGSLTGENADIKGNLTISESATFKGTRDTLTNITGN 481
QY 923 FTNNGTAEINITQGVYKLGWNTDGLNITTHAKRNORSITIGDIIINKKGSNLITDSND 982
DB 482 FTNNGTAEINITQGVYKLGWNTDGLNITTHAKRNORSITIGDIIINKKGSNLITDSND 541
QY 983 AEIOIGNISOKENLTISSDKINIKQITIKKIGDGEDSSSDATSNANITIKTKELKT 1042
DB 542 AEIOIGNISOKENLTISSDKINIKQITIKKIGDGEDSSSDATSNANITIKTKELKT 601
QY 1043 EDLSISGFNKAETAKDGRDLTIGNSNDGSGAPAKTVTFNNVKDSKISADGHNVTLNSK 1102
DB 602 EDLSISGFNKAETAKDGRDLTIGNSNDGSGAPAKTVTFNNVKDSKISADGHNVTLNSK 661
QY 1103 VKTSSNGGRESNDNDTGLTITAKNVEVNRKDTISLTKVNITASEKVTITTAGSTINATNG 1162
DB 662 VKTSSNGGRESNDNDTGLTITAKNVEVNRKDTISLTKVNITASEKVTITTAGSTINATNG 721
QY 1163 KASITTKTGDLSGTISGNTVSATVDLTTKSGSKIEAKSGEANVTSATCTIGGTISGNT 1222
DB 722 KASITTKTGDLSGTISGNTVSATVDLTTKSGSKIEAKSGEANVTSATCTIGGTISGNT 781
QY 1223 NVVTANAGDLTVNGAEINATEGAATLTATCNTLTITAGSSITSTKGOVDLLAONGSIAG 1282
DB 782 NVVTANAGDLTVNGAEINATEGAATLTATCNTLTITAGSSITSTKGOVDLLAONGSIAG 841
QY 1283 SINAANVTLTNTGTLTTVAGSDIKATSGTLVINAKDAKLGNDASGDSTEVNAVNASGSGS 1342
```

```
|||||
DB 842 SINAANVTLTNTGTLTTVAGSDIKATSGTLVINAKDAKLGNDASGDSTEVNAVNASGSGS 901
QY 1343 VTAATSSVNTITGDLNLTNGLNIISKDGRNTRVLRGKEIEVKYIQPGVASVEVIEAKRV 1402
DB 902 VTAATSSVNTITGDLNLTNGLNIISKDGRNTRVLRGKEIEVKYIQPGVASVEVIEAKRV 961
QY 1403 LEKVKDLSDDEERTETLAKLGVSAVRFEVPEPNNITITVNTQNEFTTRPSSQVITSEKACFSSG 1462
DB 962 LEKVKDLSDDEERTETLAKLGVSAVRFEVPEPNNITITVNTQNEFTTRPSSQVITSEKACFSSG 1021
QY 1463 NGARVCTNVADGGOP 1477
DB 1022 NGARVCTNVADGGOP 1036

RESULT 11
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-682-2

Query Match      67.1%; Score 4967; DB 1; Length 1536;
Best Local Similarity 67.3%; Pred. No. 9.3e-275;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRLHAKLPKLSAMLLSLGVT 60
DB 1 MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRLHAKLPKLSAMLLSLGVT 60
QY 61 SIPOSVLASGLOGMDVYHGTATMQVDGNKTIIRNSVDIINMKQFNIDQNMVQFLOENN 120
DB 61 SIPOSVLASGLOGMDVYHGTATMQVDGNKTIIRNSVDIINMKQFNIDQNMVQFLOENN 120
QY 121 NSAVFNKRVTSNQTISQKGLDLSNGQVFLNPNGTITGKDAINTNGFTASTLIDISNENIK 180
DB 121 NSAVFNKRVTSNQTISQKGLDLSNGQVFLNPNGTITGKDAINTNGFTASTLIDISNENIK 180
```

Db 121 NSAVFNRVTSNQISQLGILDSNGQVFLINPNQIITGKDAIINTNGFTASTLDISNENIK 180  
QY 181 ARNFTPEQTKD KALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGOKIT 240  
Db 181 ARNFTPEQTKD KALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSIISLLAGOKIT 240  
QY 241 ISDIINPTIYSTAAPENEAIVNGLDIFAKGGINIVRAATIRNOGKLSADSVSKDGSNIV 300  
Db 241 ISDIINPTIYSTAAPENEAIVNGLDIFAKGGINIVRAATIRNOGKLSADSVSKDGSNIV 300  
QY 301 LSAKEGEAIEGIVISAQNOQAAGKGLMITGDKVTLKTGAVIDLSKGEGETVLGDERGE 360  
Db 301 LSAKEGEAIEGIVISAQNOQAAGKGLMITGDKVTLKTGAVIDLSKGEGETVLGDERGE 360  
QY 361 GKNGIOLAKTISLEKGSTINVSKEKGGFAIVGDIALIDGNNIAQSGSDIAKTGGFVET 420  
Db 361 GKNGIOLAKTISLEKGSTINVSKEKGGFAIVGDIALIDGNNIAQSGSDIAKTGGFVET 420  
QY 421 SGHDLFIKDAIVDAKEWLLDFDNYSINAEPLFNNTGINDFPTGTG-EASDPKKNSEL 479  
Db 421 SGHDLFIKDAIVDAKEWLLDFDNYSINAEPLFNNTGINDFPTGTG-EASDPKKNSEL 479  
QY 480 KTLTNTTISNVLKNAWTHNITASRKLTVNSINSGSHLILHSKQGGVQVQIDGDI 539  
Db 479 KTLTNTTISNVLKNAWTHNITASRKLTVNSINSGSHLILHSKQGGVQVQIDGDI 537  
QY 540 ----SKGGLTIYSGGWVDVHKNIITLD--QGFLNITA-ASVAFEGGNKARDAANAKIVAQ 593  
Db 538 TGDDTRGANLTIYSGGWVDVHKNIITLD--QGFLNITA-ASVAFEGGNKARDAANAKIVAQ 593  
QY 594 GTVITIGEGKFRANVNSLGTGKGLNITISSVNN---LTHNLSTINISGNITINOTTRK 650  
Db 591 GTIT--SGNOKGFRFNNSLGTGSGLOQTTRTKNYAITNKECTNLISGKVNLSMVLPK 649  
QY 651 NTS-YWOTSHDSHWNVSALENLETCANETTF--IKYISSNSKGLITQYRSAGVNFVNGVNM 708  
Db 650 NESGYDEKFGRTYWNLSNLSVSEGEENLTIDSRGSDSAGTLTQPYNLNGISF---NKDT 706  
QY 709 SFNUEGAKVNFKLKPNENMNTSKPLP--IRFLANITATGGSGVFFDIYANHS--RGAE 764  
Db 707 TFNVERNARVNFIDKAPIGINKYSLNYSVSPNGNISVSGGSDVFTLLASSNVQTPGVV 766  
QY 765 LKMEISINSGANFTLASHVRGDDAFKINKDLTINANSNFSLRQTKDDPYDGYARNAIN 824  
Db 767 INSKYFNVSTGSSRFTYSGTKTGFSEKDLTLNATGGNITLLOVEGT--DGMIGKGIV 824  
QY 825 STYNISLGGNVTLGGQSSSITGNITIEKAANVTLEANNAPNOQIRDRVIKLGLSLV 884  
Db 825 AKKNITFEGGNITFGSRKAVTEIEGNVTINNANVTILGSDFDNHQ--KPLTIKKDVIIN 892  
QY 885 NGSLSLGENADIKGLNLTISESATFGKTRDTLNTGNTNNGTAEINITQGVVYKLGNYT 944  
Db 883 SGNLTAGNVITAGNLTUVESNANFAITNFTFNVGGLFDNKGNSNISIAKGGAREFKID 942  
QY 945 NQDGLNITTHAKNRQSIIGCDIINKGSLNITSDNNDIAEIQGGSISOKEGNLTISDK 1004  
Db 943 NSKNLSITTSSTYRTIISGNTNKGDLNITNEDSGDEMQJGGDVDSQKEGNLTISDK 1002  
QY 1005 INTKQITIKGIDGESSDATSANLTIKTKEKLTDLTSLGPNKAEITAKGDRDIT 1064  
Db 1003 INTKQITIKAGVDGSDSDATNANLTIKTKEKLTDLTSLGPNKAEITAKGDRDIT 1062  
QY 1065 IGNSDNGSCAEAKTVTFNNVNDKSLISADGHVNTLNSKVYTSNNGGRESNDNDTGLTI 1124  
Db 1063 IGNTNSAD--GTNAKKVTFNOVKDSKLSADGHVNTLNSKVYTSNNGGRESNDNDTGLTI 1121  
QY 1125 TAKNVNNDKITSKTVNITA--SEKVTITAGTATINATNCKASITTKT----- 1170  
Db 1122 DAKNVVNNNITSHKAVSISATSGEITTKTGTINATGNVEITAQTSILGIESSGS 1181  
QY 1171 -----GDISGTISGNT 1181  
Db 1182 VTLTATEGALAVSNISGNTVTVTANSALTTLTLAGSTIKGTESVTTSSQSGDITGIGT 1241

QY 1182 VSVSATVDLTTSKSGKIEAKSGEANVTSATCTIGTISGNTVNTANAGDLTVNGAEIN 1241  
Db 1242 VEVKATLTTQSSNKKIKATTGEANVTSATCTIGTISGNTVNTANAGDLTVNGAEIN 1301  
QY 1242 ATEGAATLTATGNTLTTEAGSSITSTKQVOLLAQNGSIAGSINAANVTLMNTGTLTVA 1301  
Db 1302 ATEGAATLTTSSGKLTTEASSHITSAGQVNLQAQDGSVAGSINAANVTLMNTGTLTVA 1361  
QY 1302 GSDIKATSGTLVINAADKALMGDASGDSTEYNAVNASGSGSVTAATSSSVNITGDLANTN 1361  
Db 1362 GSNINATSGTLVINAADKALMGDASGDSTEYNAVNASGSGSVTAATSSSVNITGDLANTN 1421  
QY 1362 GLNIIKSDGRNTVPLRKEIEVKYIQGVASVEVIEAKRVLEKVKDLSDEERETLAKLG 1421  
Db 1422 GLNIIKSRNGINTVLKGVKIDVKYIQGVASVEVIEAKRVLEKVKDLSDEERETLAKLG 1481  
QY 1422 VSAVRFEPPNNTITVNTQNEFTTRPSSQVSIIECKACFSSGNGARVCTNVADDO 1476  
Db 1482 VSAVRFEPPNNTITVNTQNEFTTRPSSQVSIIECKACFSSGNGARVCTNVADDO 1536

## RESULT 12

US-08-302-832-2

; Sequence 2, Application US/08302832

; Patent No. 5603938

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/302,832

; FILING DATE: 16-SEP-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9205704.1

; FILING DATE: 16-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US pct/us93/02166

; FILING DATE: 16-MAR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Berkstresser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-404

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1536 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-302-832-2

Query Match

Best Local Similarity 67.1%; Score 4967; DB 1; Length 1536;

Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

QY 1 MNKYRLKFKRLNALVAVSELARGCHDHSKSEKSEKPMKVRHLALPLSAMLGLVGT 60  
DB 1 MNKYRLKFKRLNALVAVSELARGCHDHSKSEKSEKPMKVRHLALPLSAMLGLVGT 60  
QY 61 SIQSVLASGLQGMVHVHGTATMOVDCNKTIIIRNSVDALINNKQFIDQNMVQFLOQNN 120  
DB 61 SIQSVLASGLQGMVHVHGTATMOVDCNKTIIIRNSVDALINNKQFIDQNMVQFLOQNN 120  
QY 121 NSAVFNVTNOISOLKGLDLSNGOVFLINPGLITIGKDAIINTNGFTASTLDSINENIK 180  
DB 121 NSAVFNVTNOISOLKGLDLSNGOVFLINPGLITIGKDAIINTNGFTASTLDSINENIK 180  
QY 181 ARNFTFQTKDKALAEIVNHLITVYKDGSVNLIGKVKNEGIVSYNGSISILLAGOKIT 240  
DB 181 ARNFTFQTKDKALAEIVNHLITVYKDGSVNLIGKVKNEGIVSYNGSISILLAGOKIT 240  
QY 241 ISDIINPTIYSAAPNEAVNLGDIIPAKGGINVRAATTIRNOGKLSADSVSKDKSGNTV 300  
DB 241 ISDIINPTIYSAAPNEAVNLGDIIPAKGGINVRAATTIRNOGKLSADSVSKDKSGNTV 300  
QY 301 LSAKEGEAEIGGVISAQNOQAKGKMLITGDKVTLTKGAVIDLSGKEGGTYLGGDERGE 360  
DB 301 LSAKEGEAEIGGVISAQNOQAKGKMLITGDKVTLTKGAVIDLSGKEGGTYLGGDERGE 360  
QY 361 KGNGIOLAKKTSLEKGGTINVSQKKEGFAIVWGDIALIDGNINAQSGDIATKGGFVET 420  
DB 361 KGNGIOLAKKTSLEKGGTINVSQKKEGFAIVWGDIALIDGNINAQSGDIATKGGFVET 420  
QY 421 SGHDLFIKDNAIVDAKEMWLLDFDNVSIINAEPLFNNTGGINDEFTGTG-EASDPKKNSEL 479  
DB 421 SGHDLFIKDNAIVDAKEMWLLDFDNVSIINAEPLFNNTGGINDEFTGTG-EASDPKKNSEL 479  
QY 480 KTLTNTTISVNLKAWPMNTTASRKLTVNSSINIGSNLSHLHSKQGGQVQIDGIT 539  
DB 479 KTLTNTTISVNLKAWPMNTTASRKLTVNSSINIGSNLSHLHSKQGGQVQIDGIT 537  
QY 540 ---SKGGLNLTISGQWVHVHKNITLD-QGFLNITA-ASVAFEGGNNKARDAANAKTVAQ 593  
DB 538 TGDTRGANLTISGQWVHVHKNITLD-QGFLNITA-ASVAFEGGNNKARDAANAKTVAQ 590  
QY 594 GTVITGEGKOPRANNSVINGTGLNLTSSVNN--LTHNLSTINTSGNITINOTRK 650  
DB 591 GTIT-SGNOKGRFNNVSLNGTSGLOFTTKTKYATNFEGTLNLSKGVNLSMWLPK 649  
QY 651 NTS-YWQTSHDHSHVNSALNLETGANFTF-IKYSNSKGLTQYRSAGVNFVNGVNM 708  
DB 650 NESGYDKFKRTYWNLTSLNVSSEGEFNLIDSRGSDSAGTLTQPYNLGSLF---NKDT 706  
QY 709 SFNLKEGAKVNFELKPNENMNTSKPLP-IRFLANITATGGGVFFDIYANHSQ---RGAE 764  
DB 707 TFNVNARVNFIDKAPIGINKYSSLVNYSFNGNITSVSGGSDVFTLLASSNNVOTPGV 766  
QY 765 LKMSINISNGANFTLNSHVGRDDAFKINKOLITINATSNFSLRQTKDFDYDGYARNAIN 824  
DB 767 INSKYFNVSTGSLRFKTSKGTGTSIEKDLTLNATGNITLLQVEGT--DGMITGKGV 824  
QY 825 STYNISILGNFTLGGONSSSITGNITIEKAANYTLEANNAPNOQNRDRVILKGLSLLV 884  
DB 825 AKKNITFEFGNITFGSKAVTELEGNVTINNNANVTNLGSDFDNHQ--KPLTKDKDVIIN 882  
QY 885 NGLSLTGENADTKGLNLTSEATPKGKTRDTLNTGNFTNNGTAEINTQGVVKGNTV 944  
DB 883 SGNLTAGNVINLGNLTVESNANFKATNFTFNVGGLFDKNGSNISIAKGGARFKDID 942  
QY 945 NQDGLNITHAKNRQISITGGDIINKKGSINTSDSNDAEIOIGNISOKEGLNLTSSDK 1004  
DB 943 NSKNLSITNNSSTYKTLISGNTNKGDLNITNEGSDTEMQIGGDVSOKEGLNLTSSDK 1002  
QY 1005 INTTKQITIKKGDGDESSSDATNSANLTKTKELKLTEDLSISGPNKAEITAKGDRDIT 1064  
DB 1003 INITKQITIKAGVDGNSDSDATNNANLTKTKELKLTQDLNLSGPNKAEITAKGDSULT 1062

QY 1065 IGNSNDGSGAEAKTYTFNNVYKDSKISADGHNVLNLSKVKTSSSNGGRESNDNDGLTI 1124  
DB 1063 IGNTNSAD-GYNKAKVTENQVRKDSKISADGHNVLNLSKVKTSSSNNETEDSDNNAGLTI 1121  
QY 1125 TAKNVEVNDKDTSLKTVNITA-SEKVTITAGSTINATNGKASITTKT----- 1170  
DB 1122 DAKNVTNNIIFSHKAVISATSATSGEITTKGTITINATGNVEITAGTOSILGGIESSGS 1181  
QY 1171 -----GDISGTISGNT 1181  
DB 1182 VTLTATEGALVSNISGNTVTVTANSALTTLAGSTIKGTESVTTSSGSDIGGTISG 1241  
QY 1182 VSVSATVDLTTRKSGKIEAKGEANVTSATGTIGGTIGTISGNTVNTANAGDLTVNGAEIN 1241  
DB 1242 VEVKATESLTQOSKIKATTGEANVTSATGTIGGTIGTISGNTVNTANAGDLTVNGAEIN 1301  
QY 1242 ATEGAATLTAGTTLTTEAGSSITTKGQVLLAQAQNGSTAGSINAANVTTLNTTGLTTVA 1301  
DB 1302 ATEGAATLTSSGKLTTEASSHITSAGQVNLSDAQDSVAGSINAANVTTLNTTGLTTVK 1361  
QY 1302 GSDIKATSGTLVINAKDNLGNDASGDSDEVNANVNASGSGSVTAATSSSVNTIGDNLTVN 1361  
DB 1362 GSNINATSGTLVINAKDAELNGAALGNHTVNTNANGSGSVIATTSRNVITGDLITIN 1421  
QY 1362 GLNITSKGRNTVRKGEIEVKYIQPGVASVEEVIEAKRVLEKVKDLSDDEERETLAKIG 1421  
DB 1422 GLNLSKGNLTVLLKGVKIDVYIQPGIASVDEVIEAKRVLEKVKDLSDDEERETLAKIG 1481  
QY 1422 VSARVFNENNTTVTQNEFTTRPSSOVTISEKACFSGNGCARVCTNVADGQ 1476  
DB 1482 VSARFIEPNNTITVDQNEFATRLPSRIVISEGRACFNSDGAATVCVNIADNGR 1536  
RESULT 13  
US-08-530-198-2  
; Sequence 2, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,198  
; FILING DATE: 13-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: JWB-1186  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear



i MOLECULE TYPE: protein  
US-08-530-198-2

Query Match 67.1%; Score 4967; DB 2; Length 1536;  
Best Local Similarity 67.3%; Pred. No. 9.3e-275;  
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

```
QY 1 MNKYRLKFKSRNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSALLSLGVT 60
DB 1 MNKYRLKFKSRNALVAVSELARGCDHSTKGESEKPARMKVRHLAKPLSALLSLGVT 60
QY 61 SIPOSVLASGLQGMVHVGTATQVQDGNKTIIRNSVDAILNWKQFNIDQNMVQFLQENN 120
DB 61 SIPOSVLASGLQGMVHVGTATQVQDGNKTIIRNSVDAILNWKQFNIDQNMVQFLQENN 120
QY 121 NSAVFNRTVNSQLKGIIDSGOVFLINPNGITIGDKDAIINTNGFTASTLDSINENIK 180
DB 121 NSAVFNRTVNSQLKGIIDSGOVFLINPNGITIGDKDAIINTNGFTASTLDSINENIK 180
QY 181 ARNFTPEQTKDKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSGISLLAQKIT 240
DB 181 ARNFTPEQTKDKALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGSGISLLAQKIT 240
QY 241 ISDIINPTIYTAAPENEAVALNGLDIFAKGGINVRAAATIRNOGKLSADSVSKDKSGNIV 300
DB 241 ISDIINPTIYTAAPENEAVALNGLDIFAKGGINVRAAATIRNOGKLSADSVSKDKSGNIV 300
QY 301 LSAKEGEAEITGGVISAQNOOAKGKLMITGDKVTLKTGAVIDLSKGEGETYLGDERGE 360
DB 301 LSAKEGEAEITGGVISAQNOOAKGKLMITGDKVTLKTGAVIDLSKGEGETYLGDERGE 360
QY 361 GKNGIOLAKKTSLEKSTINVSKEKGGFAIYVGDIALDGNINAQSGDIAKTGGFVET 420
DB 361 GKNGIOLAKKTSLEKSTINVSKEKGGFAIYVGDIALDGNINAQSGDIAKTGGFVET 420
QY 421 SGHDLFIKONAIVDAKEWLDFDNVSNIAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479
DB 421 SGHDLFIKONAIVDAKEWLDFDNVSNIAEDPLFNNTGINDPEFTGTG-EASDPKKNSEL 479
QY 480 KTTLTNTTISNLYKNAWTMNITASKRLTVNSSINIGSNLSHLTHSKGQGGVQIDGIT 539
DB 479 KTTLTNTTISNLYKNAWTMNITASKRLTVNSSINIGSNLSHLTHSKGQGGVQIDGIT 539
QY 540 ----SKGGLNLTISGWDVHKNITLD-QGFLNITA-ASVAFEGGNNKARDAANAIVAQ 593
DB 538 TGDTRGANLTISGWDVHKNITLD-QGFLNITA-ASVAFEGGNNKARDAANAIVAQ 593
QY 594 GTVTITGEGKDFRANVSLNGTGKGLNISSVNN---LTHNLSTGNTINSGNITNQTTRK 650
DB 591 GNTT-SGNOKGFRFNVSLENGTGSGLOQFTTTRNKYATNKFEGTLNLSGKVNISWVLPK 649
QY 651 NTS-YWQTSQSHSHWVSALNLETGANFTF-IKYISSNSKGLTQYRSSAGVNFNGVNGM 708
DB 650 NESGYDKFKRTYWNLTSLNVSSEGEFNLIDRSOSAGTLTQPNYLNGLISF---NKDT 706
QY 709 SPNLKGAQVNFPLKPNENMTSPKLP-IRFLANITATGGSGVFFDIYANHSQ---RCAE 764
DB 707 TFNVERNARVNFDPKAPTGINKYSSLAYASFNISVSGGSGVDFTLASSNVQTPGVV 766
QY 765 LKMSINISNGANFTLNHVRDGAFAKINKDLTINATNSNFSLPOTKDDFYDGYARNAIN 824
DB 767 INSKYFNVSSTGSLRFKTSKGTGTFSTEKDLTUNATGGNITLLQVEGT--DGMIGKIV 824
QY 825 STYNISILGGNVTGGONSSSITGNTIEKAAVNTLEANNAPNOQNRDRVTKLGLSLV 884
DB 825 AKKNIITEGGNITFGSRKAVTEIEGNVTINNANVTLLIGDSDNHQ--KPLTIKQVNIIN 882
QY 885 NCSLSLTGENADIKGNLTISATFETKCTROTLLNITNGFTNNGTAENITQGVVKNLGNVT 944
DB 883 SGNITAGNIVNIAGNLTIVSNANFKAITNFTNVGGLFDNKNISIAKGGAREFKID 942
QY 945 NQDGLNITTHAKRNQRSIIIGDIIINKGSLNITDSDNDAEIQGNGISQKEGNLTSSDK 1004
DB 945 NQDGLNITTHAKRNQRSIIIGDIIINKGSLNITDSDNDAEIQGNGISQKEGNLTSSDK 1004
```

```
DB 943 NSKNLSITTNSSSTYRTIISGNITNKNGLNITNEGSDTEMOIGGDVSOKEGNLTISSDK 1002
QY 1005 INITKOITTKKIDGDBDSSDSDATSNANLTIKTKEKLTEDLSISGFNKAETAKORDLT 1064
DB 1003 INITKOITTKKIDGDBDSSDSDATSNANLTIKTKEKLTEDLSISGFNKAETAKORDLT 1062
QY 1065 IGNSNDGNSGAEAKVTTFNNVKDSKISADCHNVTLSNVKVTSSNGGREGSNDTGLTI 1124
DB 1063 IGNTNSAD-GTNVAKKVTTFNNVKDSKISADCHNVTLSNVKVTSSNGGREGSNDTGLTI 1121
QY 1125 TAKNVEVNDKIDTSLKTVNITA-SEKVTITAGSTINATNGKASITTKT----- 1170
DB 1122 DAKNVTNNTTSHKAVSISATSGEITTKGTITNATNGVEITATQTSILGIESSSGS 1181
QY 1171 -----GDISGTISGNT 1181
DB 1182 VTLTATEGALAVNSITSGNVTVTANSALTLTLAGSTIKGTESVTTSSQSGDITGSGT 1241
QY 1182 VSVSATVLTITKSGSKIEAKGCEANVTSATGTIGTITSGNTVNTVNTANAGDLTVNGABIN 1241
DB 1242 VEVKATESLITQSNKIKATTEGEANVTSATGTIGTITSGNTVNTVNTANAGDLTVNGABIN 1301
QY 1242 ATEGAATITATGNTLTTEAGSSITSTKQVDLLAQAQSIAGSINAANVTLTGTLTVA 1301
DB 1302 ATEGAATITATGNTLTTEAGSSITSTKQVDLLAQAQSIAGSINAANVTLTGTLTVA 1361
QY 1302 GSDIKATSGTLVINAKDAKLANGDASGDSSTEVYNAVNASGSGSVTAATSSVNTITGLNTVN 1361
DB 1362 GSNINATSGTLVINAKDAKLANGDASGDSSTEVYNAVNASGSGSVTAATSSVNTITGLNTVN 1421
QY 1362 GLNITISKDRNTVRLRGKIEVKYIQPCVASVEBIEAKRVLEKVKDLSDEERETLAKLG 1421
DB 1422 GLNITISKDRNTVRLRGKIEVKYIQPCVASVEBIEAKRVLEKVKDLSDEERETLAKLG 1481
QY 1422 VSAVREVEPNNTITVNTQNETTRPPSOVLISEKACFSSGNGARVCTNVADQ 1476
DB 1482 VSAVREVEPNNTITVNTQNETTRPPSOVLISEKACFSSGNGARVCTNVADQ 1536
```

## RESULT 14

US-08-469-880-2  
Sequence 2, Application US/08459880  
Patent No. 5876733  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J.  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza ;  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,880  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832

```

; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-469-880-2

Query Match 67.1%; Score 4967; DB 2; Length 1536;
Best Local Similarity 67.3%; Pred. No. 9.3e-275;
Matches 1047; Conservative 146; Mismatches 264; Indels 98; Gaps 20;

Qy 1 MNKIYRLKFKSRKLNALVAVSELARGCDHSTEGSEKPKRMKVRLHAKLPKLSAMLLSLGVT 60
Db 1 MNKIYRLKFKSRKLNALVAVSELARGCDHSTEGSEKPKRMKVRLHAKLPKLSAMLLSLGVT 60
Qy 61 SIPOSVLASGLQGMVYVHGATQVGDGKNTIIRNSVDAILNKKQFIDQNMVQFLOENN 120
Db 61 SIPOSVLASGLQGMVYVHGATQVGDGKNTIIRNSVDAILNKKQFIDQNMVQFLOENN 120
Qy 121 NSAVFNRTSNOISOLKGLDNGOVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
Db 121 NSAVFNRTSNOISOLKGLDNGOVFLINPNGITIGKDAIINTNGFTASTLDSINENIK 180
Qy 181 ARNTEFQTKDCAIAEVNHLITVKGDSVNLIGKVGKNEGIVSVNGGSIISLLAGQKIT 240
Db 181 ARNTEFQTKDCAIAEVNHLITVKGDSVNLIGKVGKNEGIVSVNGGSIISLLAGQKIT 240
Qy 241 ISDIINPTITYSIAAPENEAVALGDIKAGKGNINVRATIRNQGKLSADSVSKDKSNIV 300
Db 241 ISDIINPTITYSIAAPENEAVALGDIKAGKGNINVRATIRNQGKLSADSVSKDKSNIV 300
Qy 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKGAVIDLSEKGGEGTYLGGDERGE 360
Db 301 LSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKGAVIDLSEKGGEGTYLGGDERGE 360
Qy 361 GKNGIOLAKKTSLEKGGSTINVSKEKGFALVWGDIALIDGNINAQSGDIAKTGGFVET 420
Db 361 GKNGIOLAKKTSLEKGGSTINVSKEKGFALVWGDIALIDGNINAQSGDIAKTGGFVET 420
Qy 421 SGHDLFIKDNAIVDAKEWLLDFDNVSIINAEDPLFNNTGGINDEFTGTG-EASDPKKWSEL 479
Db 421 SGHDLFIKDNAIVDAKEWLLDFDNVSIINAEDPLFNNTGGINDEFTGTG-EASDPKKWSEL 479
Qy 480 KTTLTNTTISNVLKNAWMTNITA SKITVNSSINIGSNHSLILHSKQGGGVQIDGIT 539
Db 479 KTTLTNTTLESILKGTETVNTAORIVYVNSSINL-SNGSLTLWSEGRSGGVETINDIT 537
Qy 540 ----SKGNLNTIYSGGWVDVHKHNTLD-QGFLNITA-ASVAFEGGNKARDAAAKIVAQ 593
Db 538 TGDTRGANLTIYSGGWVDVHKHNTLD-QGFLNITA-ASVAFEGGNKARDAAAKIVAQ 590
Qy 594 GTVTLTGKDFRANVSLNGTGKGLNISSVNN--LTHNLSTGNTINISGNITINOTRK 650
Db 591 GHIIT-SGNOKGFRFNNSVSLNGTSGLOQTFRKNTKAYATNFEGLTILSGKVNISMVLPK 649
Qy 651 NTS-TWQTSHDHSHVNSALNLETGANFTF-IKYSNSKGLTQYRSAGVNGVNCVNM 708
Db 650 NESGVDKFKGRTYWNLSUNVSESEFNLTIDSRGSDSAGILTQPYNLNGISF---NKDT 706
Qy 709 SFNLKEGAKVNFKLKPNENMTSKPL-IRFLANTATATGGGVFFDIYANHSG---RGA 764
Db 707 TENVERNARVNDIKAPINGINKYSSINLVSFNGSVSGGSDVPTLLSSNNVOTPGV 766

```

```

Qy 765 LKMSSEINISGANFTLNSHVRGDDAFKINKDLTINATNSNFSLQTKDDFYDGVARNAIN 824
Db 767 INSKYFNVTSGSLRFTKSTGTFSEKDLTLNATGNNITLLOVET--DGMIGKIV 824
Qy 825 STYNIISILGNNVTILGGQSSSITCNITIEKAANVTLEANNAPNOQNIIRDRVILKGLSLLV 884
Db 825 AKKNITFEKGNITFGSKRAVTEIEGNTVINNNANVTILGSDFDNHO--KPLTIKKDVIIN 882
Qy 885 NGSILGTENADIKGNLTISESATFKGKTRDTLNTITNGTAEINITGVVVKLGNV 944
Db 883 SGNLTAGNIVNIAGNLTVESNANFKAITNFTFNVGGLFDNKGNSNISIAKGARFKDID 942
Qy 945 NDGDLNITTHAKRNORSIIIGDIIKNGSLNITDNNDAEIIQIGNISQKGNLTISSDK 1004
Db 943 NKNLSITNNSSTYRTIISGNITNKGDLNITNBSGSDTEMQIGDVSQKGNLTISSDK 1002
Qy 1005 INITKOITIKKIDGEDSSDATSNANLTIKTKELKLTEDLSISGFNKAETITAKDGRDLT 1064
Db 1003 INITKOITIKAGVDGSDSDATNANLTIKTKELKLTQDLNLSGFNKAETITAKDGSDLT 1062
Qy 1065 IGNSDNGSGAEAKVTNNYKDKSKISADGHNVTLNSKVKTSSSNGGREGSDNDTGLTI 1124
Db 1063 IGNTISAD-GTNNAKVTFNQVADSKISADGHKVTILHSKVEISGSNNNTEDSDNNAGLTI 1121
Qy 1125 TAKNVEVNKDITSLKVTVNITA-SEKVTYTAGSTINATNGKASITTKT----- 1170
Db 1122 DARNVTNNNITSHKAVSISATSGEITTKGTITNATNGVEITAGTGSILGGIESSGS 1181
Qy 1171 -----GDISGTISGNT 1181
Db 1182 VTLTATEGALAVNSISGNTVTVTANGALTLAGSTIKGTESVTTSSQSGDIGTISGNT 1241
Qy 1182 VVSATFVDLTITKSGKIEAKSGEANYTSATGTTGGTISGNTVNTVNTANGADLPVNGAEIN 1241
Db 1242 VEKATPESITTSOSKIKATGEANYTSATGTTGGTISGNTVNTVNTANGADLPVNGAEIN 1301
Qy 1242 ATEGAATLTATNTLTTEAGSSITSTKGQVLLAQNGSTAGSINAAVNTLTGTLTVA 1301
Db 1302 ATEGAATLTSSGKLTTEASSHTTSAGQVNLQAQSGSVAGSINAAVNTLTGTLTVA 1361
Qy 1302 GSDIKATSTGLVINAADAKLINGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1361
Db 1362 GSNINATSGTLVINAADAKLINGDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1421
Qy 1362 GLNIISKDGRNTVRLKGEIEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEREETLAKLG 1421
Db 1422 GLNIISKNGINTVLLKGVKIDVKYIQPGIASVDVIEAKRVLEKVKDLSDEREETLAKLG 1481
Qy 1422 VSAVREVEPNNTITVNTONEFTTRPSOVITISGKCFSGSGNGARVCTVAVDGO 1476
Db 1482 VSAVREIEPNNTITVNTONEFTTRPSOVITISGKCFSGSGNGARVCTVAVDGO 1536

```

```

RESULT 15
US-08-728-470-2
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-728-470-2

```

Query Match	67.1%	Score 4967;	DB 2;	Length 1536;
Best Local Similarity	67.3%	Pred. No. 9.3e-275;		
Matches 1047; Conservative 146;	Mismatches 264;	Indels 98;	Gaps	
QY	1	MNKIYRLKFSRLNALVAVSELARGCDHSTKGEKAPARKMYRHLAKPLSAMLISLUGVT	60	
DB	1	MNKIYRLKFSRLNALVAVSELARGCDHSTKGEKAPARKMYRHLAKPLSAMLISLUGVT	60	
QY	61	SIPOSVLASGLGGDVVHGHTATQVQDGNKTIIRNSVDAIINMKQFNIDQNEWMQFLQENN	120	
DB	61	SIPOSVLASGLGGDVVHGHTATQVQDGNKTIIRNSVDAIINMKQFNIDQNEWMQFLQENN	120	
QY	121	NSAVENRVTSTNOISQLAGILDSNGQVFLNPNGLITIGKDAIINTNGFTASTDLDISLENIK	180	
DB	121	NSAVENRVTSTNOISQLAGILDSNGQVFLNPNGLITIGKDAIINTNGFTASTDLDISLENIK	180	
QY	181	ARNFTFEOTKOKALAEIVNGLITVGGKGSVNLGGKVKNEGVI SVAGGSISLLAQOKIT	240	
DB	181	ARNFTFEOTKOKALAEIVNGLITVGGKGSVNLGGKVKNEGVI SVAGGSISLLAQOKIT	240	
QY	241	ISDIINPITVTSIAAPNEAVNLGDIFAKGGINVRAATIRNOCKLSADSVSKDKSGNIV	300	
DB	241	ISDIINPITVTSIAAPNEAVNLGDIFAKGGINVRAATIRNOCKLSADSVSKDKSGNIV	300	
QY	301	LSAKEGEAETGGVTSIAONOAKGGKLMITGDKVTLTKTGAVIDLSEKGGEGTYLGGDERGE	360	
DB	301	LSAKEGEAETGGVTSIAONOAKGGKLMITGDKVTLTKTGAVIDLSEKGGEGTYLGGDERGE	360	
QY	361	GKNGTQLAKKTSLEKGSTINVSKEKGGFAI VCGDIALIDGNINNAQSGDIATGGFVET	420	
DB	361	GKNGTQLAKKTSLEKGSTINVSKEKGGFAI VCGDIALIDGNINNAQSGDIATGGFVET	420	
QY	421	SGHDLFTKDNAIVDAKEWLLDFDNVSI NAEDPLFNNTGTINDEFTPTGTG-EASDPKKNSEL	479	
DB	421	SGHDLFTKDNAIVDAKEWLLDFDNVSI NAETAGRSNTSEDDV-TGSGNSATTPXRKE-	478	
QY	480	KTTLTNTTISNLYLNKAWTNNTTASRKLTVNSNSINTGNSHLLIHSKQGGQVGDIGDIT	539	
DB	479	KTTLTNTTLESILKKTGFVNITANQRIIVYNSINL-SNGSUTLWSBGRGGGVETNNIDT	537	
QY	540	-----SKGNNITVSGGWVDVHKNTILD-QGFLNTTA-ASVAFEGGNKARDANAIAVQA	593	

Db	538	TGDDTRGANTLIYSGWVDVHKHNSILGAQGNINITAKQIDAEKGSNOV-----ITGQ	590
Qy	594	GTVTITGEGKDFRANVSLNGTGKGLNISSVNN---LTHNLSGTINISGNTINOTTRK	650
Db	591	GTIT-SQNGKGFNFVNSLNGTSGLOQTTRTKRYAITNKFEGTILNISKVNISMVLPK	649
Qy	651	NTS-YWQTSHDHWNYSALNLETGANFTF-IKYTSSNSKGLTQYKSSAGVNFVNGNM	708
Db	650	NESGYDKFKGRTYWNLTSLNVSSECEPNLTIDSRGSDSAGLTQPNLNGISF---NKDT	706
Qy	709	SNLKEGAKVNEKFLPNENWMTSPKLP-IRPLANITATGGGSVFFDIYANHSG---RGAE	764
Db	707	TFNVERNAKVNFDIKAPIGINKYISLAYFENGNISVSGGSDVFTLLASSNVQTPGV	766
Qy	765	LKMSEINISGANFTLNSHVRGDDAFKINDLTINATNSFSLROTKKDFYDGYARNAIN	824
Db	767	INSKPVNVTGSSLRFKTSGSTKTFPSIEKDLTLNATCGNITLLQVEGT--DGMIGKGIV	824
Qy	825	STYNTISLGGNVTLGGQNSSSITGNITIEKAAENVTLLEANNAPOQNIDRVIKGLSLV	884
Db	825	AKKNITFEGGNITFGSKRAVTEIEGNVTINNANVTLIGSDFDHQ-KPLTIKKDVIIN	882
Qy	885	NGSLSTGENADIKGNLTISBSATFKCTROTLTINGTNNGTAENTITQGVKLGWNT	944
Db	883	SGNLTAGNVINIAGNLTVESNAFKAITFTFNVGGLFDNKGNSNISIAKGAARFKDI	942
Qy	945	NGDNLNITHAKRNQRSIIIGDIINKGSLNITDSNNDABEIOIGGNISOKENGLTISDD	1004
Db	943	NSKLSLITNTSSSYRTIISGNTNKGDLNITNEGSDTEMOIGGDVQSKENGLTISDD	1002
Qy	1005	INITKQITIKKIGDGEUSSDATSNANLTIKTKELKLTEDLSISGFNKAETITAKDGRDL	1064
Db	1003	INITKQITIKAGVDGNSDSDATNNANLTIKTKELKLTQDLNISGFNKAETITAKDGS	1062
Qy	1065	IGNSDNGSGAEAKTVTFNVNWKDSKISADCHNVNLTNSKVKTSSSNGRESNDDTGLT	1124
Db	1063	IGNTNSAD-GTNKAKVTFNQVKDSKISADGHKVTLSHKVETSGSNWNTEDSDNAGLT	1121
Qy	1125	TAKNVENKDIITSLKTVNITA-SEKVTYTAGSTINATNGKASITKT-----	1170
Db	1122	DAKNVTYNNITSHKAVSISATSGEITTKGTITATTGNVEITAOQTSILGGIESSSGS	1181
Qy	1171	-----GDISTISGNT 1181	
Db	1182	VLTATGALAVSNI SGNTVYVTANSALITTLAGSTIKGTESVTTSSQSGDIGETISGTT	1241
Qy	1182	VSVSATVDLTKTKSGKIEAKSGEANYTSATGTIGGTISGNTVNVYTNAGDLTVGNGAEIN	1241
Db	1242	VEVKATESLTATQSNISKIKATTGEANYTSATGTIGGTISGNTVNVYTNAGDLTVGNGAEIN	1301
Qy	1242	ATEGAATLTATGNTLATEAGSSITSTKGOVDLLAONGSIAGSINAAVNTLNTTGLTTVA	1301
Db	1302	ATEGAATLTSSSKLITEASSHITAKGOVNLSDAQGSVAGSINAAVNTLNTTGLTTVK	1361
Qy	1302	GSDIKATSGTLVINAKDAXLNGDASGDSTEYNVAVNAGSGSVTAATSSSVNITGDLTNVN	1361
Db	1362	GSNINATSGTLVINAKDAELNAGALGNHTVYTNATNANGSGSVIATTSRNVNITGDLTIN	1421
Qy	1362	GLNIISKDGRNVTVRKGEIEVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLG	1421
Db	1422	GLNIISKNGINTVLLKGVKIDVKYIQPGVASVDEIEAKRILEKVKDLSDEEREALAKLG	1481
Qy	1422	VSARVFEVPENNTLTVNTQNEFTTFRSSOVIISBEGKACFPSSGNGARVCTNVADGQ	1476
Db	1482	VSARVTEPNNTLTVDTQNEFAIRPLSRIVLISEGRACFSDSGATVCVNIADNGR	1536

Search completed: March 24, 2003, 15:24:58  
Job time : 28 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:21:24 ; Search time 18 Seconds  
(without alignments)  
3403.363 Million cell updates/sec

Title: US-10-092-880-4  
Perfect score: 7407  
Sequence: 1 MNKIYLFKSKRLNALVAVS.....CFSSGNGARVCTINVADGQP 1477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	555.5	7.5	1577	1 HLYA_PROMI	P16466 proteus mir
2	521	7.0	2003	1 YDPA_ECOLI	P33666 escherichia
3	511	6.9	2249	1 OMPA_RICRI	P15921 rickettsia
4	509	6.9	917	1 HXA3_HAEIN	P45355 haemophilus
5	486.5	6.6	928	1 HXA2_HAEIN	P45354 haemophilus
6	476.5	6.4	1656	1 OMPB_RICJA	O06653 r outer mem
7	475.5	6.4	2021	1 OMPB_RICGN	O02657 rickettsia
8	469.5	6.3	3591	1 HXA3_ECORP	P12255 bordetella
9	466	6.3	905	1 HXA1_HAEIN	P44602 haemophilus
10	441.5	6.0	1325	1 YDEK_ECOLI	P32051 escherichia
11	429.5	5.8	1654	1 OMPB_RICRI	Q30471 r outer mem
12	429.5	5.8	1953	1 BIGA_SALTY	P25927 salmonella
13	428	5.8	1655	1 OMPB_RICCN	Q9Kka3 r outer mem
14	425	5.7	1608	1 HLYA_SERMA	P5320 serratia ma
15	419	5.7	1643	1 OMPB_RICPR	Q53020 r outer mem
16	415	5.6	1645	1 OMPB_RICTY	P96989 r outer mem
17	406	5.5	2660	1 YEEJ_ECO57	Q8x8v7 escherichia
18	402.5	5.4	1286	1 AIDA_ECOLI	O03155 escherichia
19	385.5	5.2	2358	1 YEEJ_ECOLI	P52143 escherichia
20	363.5	4.9	1569	1 YRJA_ECOLI	O84418 chlamydia t
21	358	4.8	1754	1 OMPB_CHLTR	P44969 haemophilus
22	357	4.8	1694	1 IGA0_HAEIN	P38058 clostridium
23	354	4.8	1672	1 OMPB_CHLMU	P45384 haemophilus
24	352	4.8	1848	1 CBPA_GLOCI	P45384 haemophilus
25	346.5	4.7	1723	1 IGA2_HAEIN	O92812 chlamydia p
26	344	4.6	1770	1 OMPB_CHLTR	O84419 chlamydia t
27	344	4.6	1770	1 OMPB_CHLTR	P14914 rickettsia
28	334.5	4.5	1300	1 L20K_RICRI	P45384 haemophilus
29	333.5	4.5	1025	1 SLAP_CAUCR	P35828 caulobacter
30	324.5	4.4	1250	1 YFAL_ECOLI	P45508 escherichia
31	322	4.3	1902	1 P2P_LACPA	Q02470 lactobacilli
32	319	4.3	1902	1 P2P_LACLC	P15293 lactococcus
33	318.5	4.3	1902	1 P3P_LACLC	P15293 lactococcus

34	317	4.3	1341	1 VG37_BPT2	P07067 bacterioph
35	315.5	4.3	1310	1 VAC3_HELPY	O48253 helicobacte
36	314	4.2	1253	1 DSSP_HUMAN	O9nz44 homo sapien
37	314	4.2	1902	1 P1P_LACLC	P16271 lactococcus
38	313	4.2	1933	1 SLAP_CAMFE	P35827 campylobact
39	308	4.2	1567	1 ICEN_XANCT	P18127 xanthomonas
40	297	4.0	1849	1 IGA4_HAEIN	P45386 haemophilus
41	296	4.0	1541	1 IGA1_HAEIN	P42782 haemophilus
42	294	4.0	1861	1 APU_THETU	P38536 t amylopull
43	290.5	3.9	1243	1 VG37_BPK3	Q38394 bacterioph
44	288	3.9	1545	1 IGA3_HAEIN	P45385 haemophilus
45	284.5	3.8	1176	1 SLAP_BACSH	P38537 bacillus sp

ALIGNMENTS

RESULT 1  
HLVA\_PROMI  
ID HLVA\_PROMI STANDARD; PRT: 1577 AA.  
AC P16466:  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Hemolysin precursor.  
GN HPMA.  
OS Proteus mirabilis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OX NCBI\_TaxID=584;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.  
RC STRAIN=Isolate 477-12;  
RX MEDLINE=90170827; PubMed=2407716;  
RA Uphoff T.S., Welch R.A.;  
RT 'Nucleotide sequencing of the proteus mirabilis calcium-independent hemolysin genes (hpma and hpmB) reveals sequence similarity with the serratia marcescens hemolysin genes (shla and shlb).';  
RT J. Bacteriol. 172:1206-1216(1990).  
RL -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.  
CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA REQUIRES HPMB FUNCTION.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA MAY BE RESPONSIBLE FOR PORE FORMATION.  
CC -1- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC  
CC EMBL: M30186; AAA25657.1; ..  
CC PIR: A35140; A35140.  
CC Hemolysin; Toxin; Outer membrane; Signal.  
CC SIGNAL 1 29  
CC FT CHAIN 30 1577 HEMOLYSIN.  
CC SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

Query Match 7.5%; Score 555.5; DB 1; Length 1577;  
Best Local Similarity 20.7%; Pred. No. 1.1e-15;  
Matches 361; Conservative 257; Mismatches 606; Indels 523; Gaps 79;  
OY 40 MKVRHLAKPLSLAMLSLGVTSIPQSVLASGL-----QGMDVW---HGTATMQV---DGN 88  
DB 1 MKSKFKLSPSGRLAASLAIFVSLNAYNGIVPDAGHQGPDSVAVNGGTQVIVITPNN 60

QY 89 KTIIRNSVDAILNKKQFNI-----DQENMVQFLOENN-----SAFNRVT 129  
Db 61 EGISHN-----QYODFNPVKGAFFNNALEAGOSQAGHLNANSNLNGQAASLIILNEV 114  
QY 130 SNOISQLK---ILDSNGQVFLNPNGITIGKDAINTNGFTASTLIDISNENIKARNFTF 186  
Db 115 SRNPSFLGQOEVFGIAAEVVLNPNGITICDGGCFINT-----SRSSLVWGNPLF 164  
QY 187 EOTKDKALAEIVNHGLITVGD-----GSYNLGGKVKNEBGVISVNGGSIISLAGQKITI 241  
Db 165 ENGOLKGYSTLNTNLLSLGKNGLNTGLDLIAPRDSRG--KITAAEISAFNGQ--TF 221  
QY 242 S---DIINPITYS-----IAAPNEAVNLGDIIPAKGNNINVAATIR 281  
Db 222 SQHFDILSSOKPVSALDSYFFGSMQSGRIIRINTAEGSGVKLAGKFTADNDLSVRADNIQ 281  
QY 282 NOGKLSADSVSKDSGNI-----VLSAKEGEAEIGGVISAQOAKGK 325  
Db 282 TDSQVRYDSYDKGSENYQYRGGITVNNSSQTLTKTELKGNITLVASHNQIKASD 341  
QY 326 LMITGDKVTLKCAVDLSKEGGETYLGDERG--EGKNGIQLAKKTSLEKGSTINYSG 383  
Db 342 LM--GDIITLQ--GADITIDGKLOQKETIDIDRWYFSWKYDVTKEKEQIQIGSQID--A 396  
QY 384 KEGGFAIVMGDIALIDGNINA-----QSGDIAKTG----- 415  
Db 397 KKNATLTATKGDVTLDAAKINAGNLA INANKDIIHNGLVEKESSENGNKRHTSRLES 456  
QY 416 -----GFVETSGHDLFIKDNAIDVAKELWLLDFDNYSI 447  
Db 457 GWSNSHOTETIKASBELTAGDLGDAOGSITPAQAKLHANENVLVNAK-----DNINL 510  
QY 448 NAEDPLFNNTGIDNDEFTGTGEASDPKKNSELKTLTNTTISNLYLKNAWTNITASRLT 507  
Db 511 NVQ--KTNDKTYTDNHVMWGGIGGQCNKNNNQOVSHAT-----QLTADGOLL 558  
QY 508 VNSSNIGCSNHLIHSKQRG-----GGVQIDGDTISKGNLTIYSGWVDV---H 556  
Db 559 LAADNNVNTIGSOV---KGNQAFVKTQGDVVDIDNALSETISKIDERTGAFLNITKSH 615  
QY 557 KNITLDQFLNITAASVAFEGGNNKARDAANAKIVAQOTVITGEGKDFRANVSLNGTG 616  
Db 616 KNET-----NK-QTSTGSELISDAQLTWS-----GNDVNVIG-- 647  
QY 617 KGLNITSSVNNLTHNLSTINI-----SGNITINQ 646  
Db 648 ---SLIKADKLGIHSLGDI NVKSAQVTKIDDEKTSIAINGHAKEVEDKQYSAGPHITH 704  
QY 647 TTRKNTSYWQTSKSHWNVSALNLETGAN--FTF-----IKYISSNS 686  
Db 705 TTNKNTS--TETEANSTISGANVDLQANKDVTFAGSDLKTITAGNASITGDNVAFVSTEN 762  
QY 687 KGLT--TOYRSSAGVNF--NGVNGNMSFNLEKAGVNF---KLKPNENMTSKPLPIRFLA 740  
Db 763 KKQTDNTDTTISGFSYTGVD-----KVGSKADQYQKHQFTQTEVTKNRGSSQTEVAG 815  
QY 741 NITATGGSVFDDIYANH--SGRGAEKMSFINISNGANFTLNSHVGRDDAFKINKDLTIN 799  
Db 816 DLTITANKDLLHEGASHHVEGRYOE--SGENIQLA--VNDSETSKTDSLNVGIDVGN 870  
QY 800 ATNSNFSLRQFKDFDGY-----ARNAINSYNITSL-----GGNVTL 838  
Db 871 LDYSGVT--KPKVKAIEDGVNVTKEGNTTDTTKKVARTADALANLANLSNLETTPNVGVEGI 929  
QY 839 GGQSSSSITGNITIEKAAN---VTEANNAPQONTDRVILKLG--SLLVN----- 885  
Db 930 KGGGSOQSOTDSQAVSTISNAGKIDIDSNKLDQGHQYQSTQEGISLTANTHTSEATLD 989  
QY 886 -----GSLISLTGEN--ADIKGNLTISESATFFKCKTRDTLNTITGNFTNNGT 928  
Db 990 KHQTFTHETGGGQIGVSTGTGSDITVAIKGEGQTTDNALMETKAGS-----QFTSNGD 1044  
QY 929 AEINITGVVVKLGNVTNDGDLNITTHAKRNQRIIGGDIINKKGSINI---TDSNDAEI 985

Db 1045 ISINVE-----NAHYEG-----AQFDAQK--GKTVINAGGDLTLAQATDTHSESQ 1089  
QY 986 QICGNISOKEGNTLTISSD-----KINTKQITIKKIDGEDSSSDATSNANLTIKT 1036  
Db 1090 NVNGSANLKVGPESKDYGGGFNAGTTHHSKEQTTAKVGTITGSQIELNAGHNLTLQ 1149  
QY 1037 KELKTELDLSGFPKAEITAKDGRDLTIGNSNDGNSGAE---KTVTFNVVKDSKISAD 1093  
Db 1150 THLSSEQDIALNATNKVDLQASSEHTEKGNLSG--GVQAGFGKKMT-----DDASSVN 1202  
QY 1094 GHNVTLNS-----KVKTSSSNGGRESNDNT--GLTITAKNVENKRDITSL----- 1138  
Db 1203 G-----LGSQAFAIGKODEKSVSRGGTINNSGNLTINGSVHLOQAQVNSKDTQLTSQ 1258  
QY 1139 -----KTVNITASEKVTITTAGSTINATNGKASITT---KT 1170  
Db 1259 DIEITSASQTDYKNNWGTDDIGFNGKKTNTPEKTEEPATSIHIGKLLLVNVEDQOKT 1318  
QY 1171 GDISGTISGNTVSYSATVDLTTKSGKIEAKSGEANTV--SATGTIGGTIS----- 1219  
Db 1319 SHONATLETGTLTINSKDLTL-----SG--ANVTADSVTGNVGGSLNIASQKESDR 1368  
QY 1220 -----GNTVNVNTANAGDL-----TVGNGAEINATEGAATLTATGN 1254  
Db 1369 HVTGVNVGVNHTNDPKSSQVNTAKAGGSLEKTIKDTIDSGIKSSTDALISOKYNSLSS 1428  
QY 1255 TLTTAEAGSITSTKGOVDLLAQNQSIAGSINAAVNTLTNTTGTTLTVAGSDIKATSGFLVI 1314  
Db 1429 TIADKTGIS--DETKAKID--OGFGKVGNGIK--NI---VTGAEGHTANADIKVT----- 1474  
QY 1315 NAKDAKLNQDASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNTVNGLNIISKDGRNTV 1374  
Db 1475 -----HVDNDADVTKTT-----SLTSSNDLSLVNNGS--TKLUGAEIVSQSQG--V 1515  
QY 1375 RLGRKELEVYKIQBPVASVEEVTEAKRVLEKVKDLSDEERETLAKLGVA--VRFVEPNNT 1433  
Db 1516 DLGGSSVKLENIE-----GHHYEAGADLDLKSSVVD-----LAKQLVGGDISFKSPVKT 1564  
QY 1434 -ITVNTQ 1439  
Db 1565 NETVNTK 1571

## RESULT 2

YDBA\_ECOLI  
ID YDBA\_ECOLI STANDARD; PRT: 2003 AA.  
AC P33666; P76087; P76088; P76856; P76857; P76859;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ydBA.  
GN YDBA OR B1401/B1405.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,





Db 1524 NVEGMNGN--NGNSSEVHYGDYTLDPVPKNTVSVTSGSDEAGSGMNLNGYVVGTVNWG 1581

Qy 1279 SIAGSINAANVTNL-----TTGTLTT-----VAGS---DIKATSGTLVINAKDAK 1320

Db 1582 S-AGKLVNNAASMGVEINTGFTAGTAGTATVSVFNVVGVNLTADAITSTSVV--WTAK 1638

Qy 1321 LMGDASGDSDEVNAVNASGSGSVTAATSSSVNITG---DLNTVGNLNIISKDGRNTVRLR 1377

Db 1639 GSTDASGNDV---VTMSKNAYTDVATDASVNDIAKALDAGYTNNELFTSLNVGTAAELN 1694

Qy 1378 GKIEVKYIQPGVASVEEVEIAKRVKLDLSEDERETLAKLGVSAREVEPENNITVN 1437

Db 1695 SALKQVS---GSOATTVFPEARVLSNRFSLMADAPKVCNGNGLAFNVVAKGDPRAELGN 1750

Qy 1438 TQNEFTT-----RPSQVSIISGKACFSSGNGARVCTNVADGG 1475

Db 1751 TEYDMLARKTIDLSQSOTMSLEYGIARLDGGAQ---RAGDNG 1791

RESULT 3

OMPA\_RICRI

ID OMPA\_RICRI STANDARD; PRT: 2249 AA.

AC P15921;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmp A).

DE OMPA.

GN Rickettsia rickettsii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_TaxID=783;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R.

RX MEDLINE=90354033; PubMed=2117568;

RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;

RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";

RL Infect. Immun. 58:2760-2769(1990).

CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A SLAYER WITH HEXAGONAL SYMMETRY.

CC -1- PTM: GLYCOSYLATED (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M31227; AAA26380.1; .

DR PIR: A41477; A41477.

DR InterPro: IPR003858; rOmpA\_rOmpB.

DR Pfam: PF02708; rOmpA\_rOmpB; 1.

DR Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.

FT SIGNAL 1 28

FT CHAIN 29 2249

FT POTENTIAL.

FT OUTER MEMBRANE PROTEIN A.

FT DOMAIN 212 1180

FT REPEAT 212 286

FT 13 X APPROXIMATE TANDEM REPEATS.

FT REPEAT 212 286

FT A (TYPE I).

FT REPEAT 287 358

FT B (TYPE II).

FT REPEAT 359 430

FT C (TYPE II).

FT REPEAT 431 505

FT D (TYPE II).

FT REPEAT 506 577

FT E (TYPE II).

FT REPEAT 578 652

FT F (TYPE I).

FT REPEAT 653 724

FT G (TYPE II).

FT REPEAT 725 799

FT H (TYPE I).

FT REPEAT 800 874

FT I (TYPE I).

FT REPEAT 875 949

FT J (TYPE I).

FT REPEAT 950 1021 K (TYPE II).

FT REPEAT 1022 1093 L (TYPE II).

FT REPEAT 1094 1165 M (TYPE II).

FT REPEAT 1166 1180 TYPE I (INCOMPLETE).

SO SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match 6.9%; Score 511; DB 1; Length 2249;

Best Local Similarity 22.1%; Pred. No. 1.1e-13;

Matches 335; Conservative 197; Mismatches 593; Indels 388; Gaps 68;

Qy 119 NNSAVENRVTSNOISQKGLDSNGQVFLINPNG-----ITIGKDAIINTNGFTAST 171

Db 47 NNAAFSNNVGNWNNNEITAAGVANG-----TPAGGPQNNWAFYTGDDYV-----T 93

Qy 172 LDISNENIKARNFTFEOTKDKALAEIVNHLITVYKDGSVNLGGKVKNEGIVSVNGSI 231

Db 94 ADAADRIIKAINVA-----GTPVPGLNITQNTVVGSIITKG-----NLLPV 134

Qy 232 SLLAGQKITIIDIINPTIVYSIAAPENAVNLGDIFAKGGNINVRATIRNOGKLSADSV 291

Db 135 TLNAGKSLTLNGNNAVAANHGFDPADPNYTGIGNIALGGAN-----AALI-----I 180

Qy 292 SKDKSGNIVLSAKEGEAEITGGVISAQNOQAKGKMLITGDKVTLKTGAVIDLSKGEGET 351

Db 181 QSAAPSKITLA---GNIDGGIITVKTDAINGTIGNTALATVNVGA-----GTA 228

Qy 352 YLGG-----DERGEGKNGIOLAKKTSLEKSGTINVSKEGKGFALVWGDIALIDGNIN 404

Db 229 TLGGAVIKATTIKLTNAASVLTLTNANAVLTGAIDNTGDNVGNLNLGALSQVTDIG 288

Qy 405 AGSGDIATGGFVETSGHDLFIKDNAIVDAKELWLDLFDNVSINAEPLFNNTGINDEFP 464

Db 289 NTNSLATISVGAGTATLG-----GAVIKATTIKLTDAASAVKFTNPVVVTAIDN--- 338

Qy 465 TGTGEASDPKKNSELKTLTLNTTISNYLKNA---WTMNTITASKRLTVN-----SSINTG 515

Db 339 ---TGNA-----NGVFTFGNTGTGNTGNVGNNTALATVNVGAG--LLQVQGVVVKANTINLT 390

Qy 516 SNSHILILSKGORGQVQIDGDIITSGK-----GNLTIYSGGVWVHVKNITLDQCF 565

Db 391 DNASAVTET-----NPVVVVTGAIDNTGNANNGIVFTGNTSV--TGDIGNTALATVNVGA 444

Qy 566 LNIITASVAFEGGNKARDANAIAVQGVITITGEGKDFRANNSVLSNLTGKG--LNIISS 624

Db 445 GTATLGGAVIKATTIKLTNAASVLTLTNANAVLTG-----AIDNTGDNVGNVNLNCA 498

Qy 625 VNNLTHLSGTINIS-----GNITINQTTKNTSYWQTSKSHSHSNVYALN--LE 671

Db 499 LSQVGTGNTGNINSLATISVGAGTATLGGAVIKATTIKLT-----DAASAVKFTNPVVV 551

Qy 672 TGANFTFIKYISSNSKGLTQYRSSAGVNFNG---VNGNMSEFLKEGAKVNEKLPNENM 728

Db 552 TGA-----IDNTGNANNGIVT-----FTGNSTVTDIG----- 579

Qy 729 NTSKPLPIRFLANI-----TATGGSVFEDLIYANHSRGAEKLMKSEIN--ISGANFTLNS 782

Db 580 NTNS-----LATISVGAGTATLGGAVIKATTIKLTNAASVLTLTNANAVLTGA---IDN 630

Qy 783 HVRGDDAFKINKDLTINATNSFSLROTDDFYDGYARNAINSTVNIILGNVTLGG-- 840

Db 631 TTGGDN-----VGVNLNLGALSQVTDIGI-----GNTNSLATISVGAGTATLGGAV 675

Qy 841 -QNSSSITGNITTEKAANVTLEANNAPNOONIRDRVIKGLSLVNGSLSLUTGENADIKG 899

Db 676 IKATTTKITNAVSARKFTN-----PVVVTGAIDSTG--NAN--NG 711

Qy 900 NLTISESAFEGKTRDTLNTITGNFTN--NGTAEINITOGVVKLGNVTDGDLNITTHAKRN 958

Db 712 IVTFTGNST-----VTGDIIGNTALATVNVGAGTATLGGAV---IKATTTKLTN 757

Qy 959 QRSIIGDDIINKKGS-----NITDSNNDAEIOIGGNISOKEGNLTISSKINIKQITI 1013

Db 758 AASVL---TLTNANAVLTGAIDNTGDNVGNLNLGALSQVTDIG-----NTNSLATI 809

CC 1014 KKGIDGESSDSSDSSNANITIKTKELKLTEDLSISGFENKAETAKDGRDLTIGNSNDG-- 1071  
 CC 1015 KKGIDGESSDSSDSSNANITIKTKELKLTEDLSISGFENKAETAKDGRDLTIGNSNDG-- 1071  
 CC 810 SVG-----AGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAVDNTGGDNVGL 862  
 CC 1072 --NSGAETVFNKVD--SKISADGHNVTLN-----SKVTSSSSNGRE 1113  
 CC 863 NLNGALSOVTDIGTGNLSLTISSVAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLT 922  
 CC 1114 SNSDNDG-----LFIITAKNVEVKKDI---TSLKTVNITASEKVTITAGSTINATNGKA 1164  
 CC 923 GADNTGTGGDNVGLNGLALSOVTDIGTGNLSLTISSVAG--TATLGGAVIKAT----- 976  
 CC 1165 SITTKGIDISGRIS--CNTVSVSATVDLTTKSGSKTEAKSGEANVTSATITGTTISGNTV 1223  
 CC 977 --TKLTDAASAVKFTNPVVTGAIDNTGNANGIVTFTGNSTV--TCNVGNTNALATV 1031  
 CC 1224 NVTANAGDLTVNGA----EINATEGAALITATGNTLTTEAGSSITSTKGOVDLLAONGS 1279  
 CC 1032 NV--GAGLLQVQGVVVKANTINLTNDASAVTFNPNVVTGAIDNTGNANGIVTFTGNST 1089  
 CC 1280 IAGSINAAN--VTLNTTGLTTVVAGSDIKATSGTLVINAKDLKNGDASGDSTEVNAVNA 1337  
 CC 1090 VTGNVGTNALTAVNVGAGLLQVQGVVVKANTINLTNDASAVTFNPNVVTGAIDNTGNA 1149  
 CC 1338 SGSGSVTAATSSSVNITGDLNTVNGL-----NIISKDGRNTVRLR- 1377  
 CC 1150 N-NGIVTFTGNSTV--TGDIGTGNALTAVNVGAGITLQAGGSLAANNIDFGARSTLEFNG 1206  
 CC 1378 -----GKEIEVKVITQGVASVEEIVIAKRVLEKVKDLSDEERTLAKLGVSAVRFEVPPN 1432  
 CC 1207 PLDGGGRAPY--YFKGAIANGNNAI--LNVNTKLTATSLTIGTVAEINIGA-----GN 1257  
 CC 1433 TIV-----NTQN--EFTTRPSSQVI-----ISEGRACFS 1460  
 CC 1258 LFTIDASVGDVTLNAQINFRARDSVLVLSNLTGTVGNVILLAAIDVAPGADEGTWVFN 1317  
 CC 1461 SG-NGARVCTNVA 1472  
 CC 1318 GGVNGLNVGSNVA 1330

## RESULT 4

HA3\_HAEIN STANDARD; PRT; 917 AA.  
 AC P45355; 1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization  
 protein A).  
 GN HXUA.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NTHI N182;  
 RX MEDLINE=95115556; PubMed=7815944;  
 RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,  
 RA Mueller-Eberhard U., Hansen E.J.;  
 RT "The 100 kDa haem:haemopexin-binding protein of Haemophilus  
 RT influenzae: structure and localization."  
 RL Mol. Microbiol. 13:863-873(1994).  
 CC -!- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL; U08349; AAA74139.1; --  
 CC TRANSPOSIT; Signal; Repeat.  
 CC SIGNAL 1 21 BY SIMILARITY  
 CC CHAIN 22 917 HEME/HEMOPEXIN-BINDING PROTEIN.  
 CC DOMAIN 24 38 3 X 5 AA TANDEM REPEATS.  
 CC REPEAT 24 28 1-1.  
 CC REPEAT 29 33 1-2.  
 CC REPEAT 34 38 1-3.  
 CC DOMAIN 111 668 6 X 6 AA APPROXIMATE REPEATS.  
 CC REPEAT 111 116 2-1.  
 CC REPEAT 203 208 2-2.  
 CC REPEAT 277 282 2-3.  
 CC REPEAT 399 404 2-4.  
 CC REPEAT 624 629 2-5.  
 CC REPEAT 663 668 2-6.  
 CC DOMAIN 159 170 2 X 6 AA APPROXIMATE TANDEM REPEATS.  
 CC REPEAT 159 164 3-1.  
 CC REPEAT 165 170 3-2.  
 CC SEQUENCE 917 AA; 100148 MW; 7FF39BB8C046539D CRC64;  
 Query Match 6.9%; Score 509; DB 1; Length 917;  
 Best Local Similarity 24.4%; Pred. No. 4.4e-14;  
 Matches 254; Conservative 167; Mismatches 372; Indels 248; Gaps 52;  
 QY 62 IPQSVL-ASGLOGMD-VVHGATMQ--VDGNKTIIRNSVDALINWKQINIDQENMVQFLQ 117  
 DB 26 LPQOGLPOOGLPOEDKVVVQATFDKTIADKMTINOTSDKVOINWHSFIDQNKVEFEKQ 85  
 QY 118 ENNSAVFNVTNSOISQIKGLDSNGOVFLNPNGITIGKDALINTNGFTASTLDIS-- 175  
 DB 86 PSENSAVYRVNFGGASQIGKLTANGKVLNPNPVIITKGAENVAGLLATTKDLQEI 145  
 QY 176 NENIKARNTFPTQDKAL---AEIVNHGLITVGDGVSNLIGKVGKVEGVSNGSIS 232  
 DB 146 SENSNSYQFT-RTKDRQVVKGVINEGEIK-AKD-FVVLNGDEVINKGNINVEKNS-- 200  
 QY 233 LLAQOKITISDIINPTITYSIAAPENAVNLGDIKAGNINVRATINQCKLSADSVS 292  
 DB 201 -TINGKVYLSSGVNFTTSL--DSGISVALEDNTVRG-----IVNDESI----- 242  
 QY 293 KDKSGNIVLSAKEGEAEI-----GGVISAQOQAGKGLMTGDKVTLTKT-----GAV 340  
 DB 243 --KAGEITLSAK--GRKEALDSLVMNNGVLEATKVSNNRNGKVLSADNVOLNNEKIKGEI 299  
 QY 341 IDLSGKGGETVYLGDERGEGKNGIOLAKKTSLEKGS-TINVSQKKEGGAIVMGDIALI 399  
 DB 300 VNF---GTEVTSNEDKK-----LKITSKTSKVTSPKINFGKSVN-----I 338  
 QY 400 DGNINAGSGDIAKTGGFVETSGHDLFIKDNAIVDAKEWLLDFDNYSINAEDPLFNNTGI 459  
 DB 339 NGNGRENSG-----THYNEERKTL-NTEV 362  
 QY 460 NDEFTPTG-EASDPKNSKLTTLTNTTISNVLKNAWTMNITASRKLTVNSSINI----- 514  
 DB 363 NIDVPAENIRIADKDNTEFDSFIQTEALSSLLANNGKVLKGN-DVNISGNINIDSFR 421  
 QY 515 GSNSHLILHSGQGGGVQID-GDITSKGNLTIYS-GGVVDVHKNTLDDQGLNLTAAAS 572  
 DB 422 GTDSSLKLTNOGH-----INIDHADINKSGRLFVVTSLQNDVDFKSNITITDSKINLNGA 477  
 QY 573 VAFEGGNKARDAANAKIVAGQVTITIGEGKDFRANKVNSLNGTGKGLNITISVNNLTHNL 632  
 DB 478 MGLGRSVNEDDYATRWK-----RAEKSQRKKFNVDMRNV-----VFNQVDEVI--L 521  
 QY 633 SG---TINISGNITINOT-----TRKNTSYWQTSKSHSNVNSALNLETGANFTFIKY 681  
 DB 522 AGGPEKVNLDKIVATGOTNEFYIDGGVSRNRSY-----EYGV----- 558  
 QY 682 ISSNSKGTITQYRSAGVNFNGVNGNMSFNLKEGAKVNFKLKPNENMNTSKPLPIRPLAN 741

```

Db 559 -----LDLRYRTQLS-----ELEQRR-RWRYRDLDLDMNKAYLYR----- 594
Qy 742 ITATGGGVSFDIYA-NHSGR----GAEKLM--SEINISNGANFTLNHVRGDDAFKINK 794
Db 595 -----FDLAKNNGSRSTKTEINISNINLNKNGFVHLLAENIKLDSN--KV 641
Qy 795 DLTINATNSFSLRQTDKDDYDGYARNAINSTYNISLTGG--NVTLGGONSSSTGNI 851
Db 642 DITFDKNSQDTLAQTNRLMGKGV-SMINS--HIKIVGDEKIGISPTGYATMFLIGEL 698
Qy 852 TIEKAA-----NVTLEANNAPNOONIRDRVILKIG-----SLLVNGSLSLT 891
Db 699 IGEKSSIFVKSHOGYTFKTDGDKIAKGN--SKEDLKITAINTGGRAAEVLLINGALSA 756
Qy 892 GENADIKG-NLTISGATPKGTRDTLNITNGTNGTAEINITQGVVKGNTVNDGDLN 950
Db 757 DNDANIANMAFTICDSANYK-TTIEVNADITAPNGGTAYLSKQVIEVKPNSFTTFFE 815
Qy 951 IITHARNQORSITGGDI-INKKGSLNITDSNNDAEIQTIGGNSIQKQEGNLTISDDKINIK 1009
Db 816 LPREKLNQTKINGASTKLSRGFARLYDKINGVR---ASNLSAQLNVTDASEKIINTK 872
Qy 1010 QITIKKKGIDGEDSSSDATSA 1030
Db 873 LVS---SLDVEKLVSVAVCD 890

RESULT 5
HX42_HAEIN
ID HX42_HAEIN STANDARD; PRT; 928 AA.
AC P45354;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
DE protein A).
GN HXA4.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-35.
RC STRAIN-DL42 / Serotype B;
RX MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.;
RT "The 100 kDa haem:haemopexin-binding protein of Haemophilus
RT influenzae: structure and localization.";
RL Mol. Microbiol. 13:863-873(1994).
RN [2]
RP SEQUENCE OF 1-30 FROM N.A.
RC STRAIN-DL42 / Serotype B;
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yorgev R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
RT heme:hemopexin by Haemophilus influenzae type b.";
RL J. Bacteriol. 177:2644-2653(1995).
CC -!- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U08348; AAA74138.1;
CC Transport; Signal; Repeat.
KW SIGNAL
FT SIGNAL
FT CHAIN 22 928 HEME/HEMOPEXIN-BINDING PROTEIN.

```

```

FT DOMAIN 101 679 6 X 6 AA APPROXIMATE REPEATS.
FT REPEAT 101 106 1-1.
FT REPEAT 205 210 1-2.
FT REPEAT 279 284 1-3.
FT REPEAT 410 415 1-4.
FT REPEAT 635 640 1-5.
FT REPEAT 674 679 1-6.
FT DOMAIN 149 172 4 X 6 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 149 154 2-1.
FT REPEAT 155 160 2-2.
FT REPEAT 161 166 2-3.
FT REPEAT 167 172 2-4.
SQ SEQUENCE 928 AA; 101228 MW; 67D45466A4B92390 CRC64;

Query Match 6.68; Score 486.5; DB 1; Length 928;
Best Local Similarity 23.8%; Pred. No. 3.7e-13;
Matches 249; Conservative 166; Mismatches 363; Indels 267; Gaps 49;

Qy 77 VHGTATMQ--VDGNKTIIRNSVDAIINWKQFNIDQNMVQFLOENNNNSAVENRVYSNQIS 134
Db 33 VFGTVTTEKTTADKMTIKQGSQDKAQIDWKSFDCQKKEVFEQNEHAYVNRVIGNAS 92
Qy 135 OLKGILDSNGOVFLINPENGITIGKDAIINNGFTASTLDIS--NENIKARNEFTFEQTK 192
Db 93 QIQGKLTKANGKVIYLANPVGIIITQGAENVAGLIATTKDLERISENSNSYQFT--RRTK 151
Qy 193 ALAEIVNHGLITVKGDSVNLIGGKVKNEGVIS-----VNGGSISLLACOKIT 240
Db 152 ---QVLKEGLVL--KDGQV-VKEGOVINEGNIITAOQDPVVLNGDEVINKGINYEKNSTIN 205
Qy 241 ISDIINPTIYSTAAPENE-AVNLGDIFAKGGININRAATIRNOGKLSADSVSKDSGNI 299
Db 206 GKVLSSGYNFTETLPDPSGISVALED-----NTVQGVKNESGI-----KAGEI 249
Qy 300 VLSAKEGEAEI-----GGVISQAQOAKGKLMITGDKVTLTKTGAVIDLSGKEGGTYL 353
Db 250 TLSAKGRKQALDSLWNNGVLEATKVSNGKGYLSADNVLNNEGNI-----KGEIVTF 304
Qy 354 GGD--ERGEKNGIQIAKKTSLTEKGS-TINVSKEKGGFAIVGWDIALIDGNINAQSGD 410
Db 305 GADVTNKKELKDNKIKTSKTSKVTSPKINFTGKSVN-----INGNFGREDS-- 351
Qy 411 IAKTGVFVTSQHDLFKONAIVDAKEWLDF-DNVSINAEDPLNFNTINDEFFTGCE 469
Db 352 -----TTHYKDEPKLNTENV-----IDVPDENIRIAD-IEDNTGTG--ITGTGT 394
Qy 470 ASDPKKNSLKTTLNTTISNVLKNAWTNITASRKLTVNSSINI-----GSMNHLILSK 525
Db 395 SS-----FIQTGALSSLLANNGKVNKGN-VNLSGRIHIDSPRGSDSLKLTKNK 443
Qy 526 GORGGVQIT-DGDITSKGNLTYS--GGWVDVHKNTITLDQGFNITAAASVAPEGGNNKAR 583
Db 444 GH-----IDINNADIHSKGRLEFFITSLQNEEDPKSNITITDSKINLNGAMGL-GRSVD 498
Qy 584 DAANAQIVAOQVTTITGEKDFRANNSLN-----GTGKGLNISSVNNLTHNLST 635
Db 499 DYDNRWQTEGS---QRKFDVKMSNVEFNQVDDVILAGGFEKVNLDKIV-----ATQG 549
Qy 636 IN--ISGNITINQTTTRKNTSYWQTSQSHSHWNVSALN-----LETGANFTFIKYI 682
Db 550 TNFYIDGGVSRN---GRKYEVGLDLDKRTQLSELNQGRRRWYGYDELDNRYAL---- 603
Qy 683 SSNSKGLTQYRSSAGVNFNGVNGMSPNLEKAGAKVNFKLPKNMNMNTSKPLPIRFLANI 742
Db 604 -----YR----- 605
Qy 743 TATGGGVFFDIYAHNSGRGAELKMSKSEINISNGANFTLNS---HVRGD---DAFKNKD 795
Db 606 -----PDLFATKKTGRSTIKDTINTSN-SNIMKNGFVHLLAEKIKLDSNKT--D 653
Qy 796 LTINATNSNFSRLQTKDDFDGYARNAINSTYNISILGGNVTL--GGQNSSSSITG---- 849
Db 654 ITFDKNSQDTLAQT-----NRLGMNGKVSIMSHIKIVGDEKIGISPTGYATM 703

```

QY 850 -----NITIEKANVTLEANNAP-----NOQNRDRVILKLG-----SLVNGS 887  
 Db 704 FLIGELIGBEKSLFVKSHOGYFKTDGNTKYIAGKYSKEDLKITAINTGGRAAEVLLINGA 763  
 QY 888 LSLTGENADIKG-NLTISSEATFKGKTRDTLNTGNTNGTAEINITQGVVKLGNTVD 946  
 Db 764 LGSADNDANIANMAFTIGDSANTK-FTIENADITAPNGNGTAYLSSKDVLEIEVPNSNF 822  
 QY 947 GDLNITHAKRNQRIIGDI-INKKSGLNITDSNNDAEIQIGNISQEGNLTISSEDKI 1005  
 Db 823 TFFELPREKLNKTINGASTKLSERGFARLYDKINGVR---ASNLSABQLNVTDASEKI 879  
 QY 1006 NITQKITIKKGDGEDSSSDATNSA 1030  
 Db 880 INTKLVS---SLDVEKLVSVAVCDA 901

## RESULT 6

OMP\_B\_RICJA STANDARD; PRT: 1656 AA.  
 AC 00653; 2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rompb)  
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
 GN OMPB.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 ON NCBI\_TaxID=35790;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YH;  
 RC Uchiyama T.;  
 RA "Sequencing of the gene encoding the protein rOmp B of Rickettsia japonica."  
 RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: AB003681; BAA20138.1;  
 CC InterPro: IPR003858; rOmpA\_rOmpB.  
 CC Pfam: PF02708; rOmpA\_rOmpB; 1.  
 CC Antigen: S-layer; Cell wall.  
 CC FT CHAIN 1 1338  
 CC FT CHAIN 1339 1656 120 KDA SURFACE-EXPOSED PROTEIN.  
 CC FT DOMAIN 528 533 32 KDA BETA PEPTIDE.  
 CC FT DOMAIN 1339 1656 POLY-GLY.  
 CC SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 6.4%; Score 476.5; DB: 1; Length 1656;  
 Best Local Similarity 21.8%; Pred. No. 1.9e-12;  
 Matches 346; Conservative 204; Mismatches 596; Indels 437; Gaps 78;

QY 34 SEKPARMKVRHLALPLSAMLISLG-----VTSIPOSVLASGLQGMVDVHGTATMQVDG-- 87

Db 2 AOKPNFLK-----KLISAGLVLTASTATIVASFAGSANGAAIQOQNRRTNGVAT--TVDGVG 54  
 QY 88 -NKTLRNSV-----DAIINKKOFN-IDONEMV-----OFLQENNSAV-----FNR 127  
 Db 55 FDOIVALANVAVANAVITANANGINLNTPAGSFENGLFSLSNANNAVTVSEDTTLGFIN 114  
 QY 128 VTSNOISOLKILDSNGOVFLINPNT-----IKDKAIIN----- 165  
 Db 115 NAANNANFNLTDR-GKTLITQGITNVOSAAHNAQINAVAKFENGGAATANDLSLG 173  
 QY 166 ---GFTASTL--DISNENIKARNFTFEQTKDKALAEIVNHGLITVKGSGVNLIGKVK 219  
 Db 174 TIDFGAAASTLVFDLAN-----PTQKAPLIILADNALIVANGANGILNTVNGFIO 222  
 QY 220 NEGVISVNGGSISLLAGOKITISLIINPTITYSIAAPENEAENVLGDIFAKGGINVRAAT 279  
 Db 223 -----VSDKSFATV-----KAINIGD---GQGFMEFNTN-AT 249  
 QY 280 IRNOGKLSADSVS-----KDKSGNIVLSAKEGEAEIGVISAQNOQAKGGLMITGDKVT 334  
 Db 250 NANALNQAGGTTINFNGTDTGRLVLLSKNGAATDFNVTGSLGNLKG-----IE 301  
 QY 335 LKTGAVI-DLSKGEGETYLGDERGEGKNGIOLAKKTSLEKGSTINVSKEGKGFAIWM 393  
 Db 302 LNTVALINGOLIANAPANAVIGTNNAGR---AAGFVVSVNDNGKAATIDGOVYAKDMVI- 357  
 QY 394 GDIALIDGINAOGSGDIATKTGFV-ETSGHDLFKDKNAIVDAKEMWLLDFDNVNSINAEPP 452  
 Db 358 -QSANANGOVNFRHIVDVGIDGTTAFKTAASIVALTQNSNFGT-----TDFGNLAAQVTP 412  
 QY 453 LENNVGINDEFPTGCEASDPKKNSELKTLTNTTISNLYKNATWMTNTTASRKLTVNSSI 512  
 Db 413 ---DTWLTGNF---TGDANNPONTAGVITFAANGILASASADA---NVAVTNTIATEAS 464  
 QY 513 NIG-----SNSHLILHSKGGGGVQI-DGDITSKGNLTIYSGWVDVHKNTLDOGFNL 567  
 Db 465 GVGCVOLSTHFAELRLGNAGSVFKLADGTVINGKVNQTVLVGG-VLAAGATLD-GSAT 522  
 QY 568 ITAASVAFEGGNKARDAANAKIVAOGVTITTEGKDFRANNVSLNG-----TGKGLN 620  
 Db 523 ITGD--IGNGGGGAALQSITLANDATKTLTGG-----ANTISANGGTINFGANGGTTK 574  
 QY 621 ITSSVNNL-----THNLSGTINI----- 638  
 Db 575 LTSTQNNIVVDCDLAIATDQTGVVDASSLTNAQTTLTISGTIGLIGANNTLGOFNIGSSK 634  
 QY 639 -----SGNITINOTTRKNTSYWQTSKSHSNVNSALNLETGANFTFIKYSISNKGSLTOYR 694  
 Db 635 TTLNGGNVALNELVIGNNGSVQFAHNTYLTITRTTNAAGOKIIFNPVNNNT-----T 687  
 QY 695 SSAGVNF-NGVNGNMSFNL-KEGAKVNEKLPKNENMTSKPLPIRELANITATGG--GSV 750  
 Db 688 LAAGTNLGSAAANPLAEINFGSKCARADTVLNVCEGYNL-----YATNITTTDANVGSF 740  
 QY 751 FFDIYAHISGGAELKMSLEINISNGANFTLNSHVGRDGAFFKINKDLTINATNSFSLROT 810  
 Db 741 VF---NAGK-----NIVSGT-----VGGQGNKFPNTVALDNGCTTVKF----- 775  
 QY 811 KDFYDGYARNALNSTVNSILGCVNLGQNSSSSITGNITIEKAA----- 857  
 Db 776 -----LGNATFN-----GNTTIAA-NSTLQISGNTADTFASADGCTGIVEFVNT 818  
 QY 858 ---NVTLEANNAPNOQNRIDRVIKLGLSVLNGSLTGENADIKGNITISE---SATFKG 911  
 Db 819 GPINVTNLKQAVP-----VNALKOITVSG-----PGNVVNEIGNAGNYHG 859  
 QY 912 KTRDTL-----NITGN-----FTNNGTAE-----INTQCG 936  
 Db 860 AMTDTIAFENSSLGAFLVPSGPFNDAGNTIPLTIKSTVGTGNETAEGFSVPVSGVDS 919  
 QY 937 VVKLGNTVNDGLNITTHAKRNQRIIG-----GDIINKK-----GSLNITDSNDA 983

Db 920 VIADQVIGD-----QNNIVGLGSDNGIIVNATTLTYAGIGTIN-----NNQG 963  
 QY 984 EIQIGNISQKGNL-----TISDKINIKTKQITIKKID--GEDSSSDATSNANLTIKT 1036  
 Db 964 TVTLSGGVPNTPTGTVGLGTIGASKF---KQVTFDDYNLNLNIIATNTINDGVTVTT 1020  
 QY 1037 KELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEKTVTFNNVSKISADGHN 1096  
 Db 1021 GGIAAG---GIAAG-----TDFDCK-ITLGSVN-GNANVRFADGIFSNSTSMIVTAKAN 1069  
 QY 1097 VTLNSKVKTSSSNGG-----RESNDDNTGTITAKNVEVKNKDJTSLKTVNITASE 1147  
 Db 1070 GTVTYLGNAFVGNIGSDPVPASVRFSGSNAGL-----KNIIVSQVIDFQTYNLGIIVN 1124  
 QY 1148 KVTYTAGSTINATNGKASTTGTGTGDISGTPIS--GNTVSVATFVDTLTKS-GSKIEAKSGE 1204  
 Db 1125 SNVILGGST-TAINGKIDLTTLTFAFGCTSGWNTSTIETTLTLANGNIGHIVIAEQAQ 1183  
 QY 1205 ANVTSATGTIGTISGNTVNVNTANA-----GDLT---VNGAEINATGAATLTATGNTL 1256  
 Db 1184 VNATT-TGT-----TTINVDNANAFSGTQYTLTIGGARFNGTLGGPNFTVT----- 1231  
 QY 1257 TTEASSITSTGQVLDLAQNGSIAGSINAANVTLTGTTLTTVAGSDIKATSGTLVINA 1316  
 Db 1232 ----GSNREVNGLIRAAANQDVVITRTNNAENIVTNDI-TNSPFGGAPGVQNVVTFVNA 1286  
 QY 1317 DKAKLNGD---ASGDSDEVNAVNASGSGSVTAATSSSVNITGDLNVTNGLNIIISKQGRNT 1373  
 Db 1287 TWTAYNALLLAKNSADSANFVGTITVTDTSAAITNAQLDVAKDIAQQLGNRLGALVILGT 1346  
 QY 1374 VRLGRKEIEVKYIOPGVASVBEVIE 1398  
 Db 1347 PEMVSGEAGA--IPAAVAAGDEAVD 1369

RESULT 7  
 ID OMPA\_RICCN STANDARD; PRT: 2021 AA.  
 AC Q52657; P95592; P95593; P95594; Q52667; Q52668; Q52669;  
 AC Q52670; Q52674;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface  
 antigen) (rOmpA) (rOmp A).  
 DE OMPA OR RC1273.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RC MEDLINE=94171067; PubMed=8125327;  
 RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;  
 RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia  
 conorii (Malish 7 strain).";  
 RL Gene 140:115-119(1994).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RC MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
 RL Science 293:2093-2098(2001).  
 [3]  
 RN SEQUENCE OF 8-204 FROM N.A.  
 RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;  
 RX MEDLINE=97015921; PubMed=8862558;  
 RA Roux V., Fournier P.E., Raoult D.;  
 RT "Differentiation of spotted fever group rickettsiae by sequencing and

analysis of restriction fragment length polymorphism of PCR-amplified  
 DNA of the gene encoding the protein rOmpA.";  
 J. Clin. Microbiol. 34:2058-2065(1996).  
 [4]  
 RN SEQUENCE OF 953-2012 FROM N.A.  
 RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;  
 RA Raoult D., Fournier P.E., Roux V.;  
 RT "Phylogenetic analysis of spotted fever group rickettsiae by study  
 of the outer surface protein rOmpA.";  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 CC FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).  
 CC !- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 CC S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC !- PTM: GLYCOSYLATED (BY SIMILARITY).  
 CC !- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; U01028; AAA17405.1; -;  
 CC EMBL; AF008674; AAL03811.1; -;  
 CC EMBL; U43794; AAB49549.1; -;  
 CC EMBL; U43798; AAB49550.1; -;  
 CC EMBL; U43806; AAB49551.1; -;  
 CC EMBL; U45244; AAB49556.1; -;  
 CC EMBL; U46918; AAB49566.1; -;  
 CC EMBL; U83440; AAC35176.1; -;  
 CC EMBL; U83443; AAC35179.1; -;  
 CC EMBL; U83448; AAC35184.1; -;  
 CC EMBL; U83453; AAC35189.1; -;  
 CC InterPro: IPR003858; rOmpA\_rOmpB.  
 CC Fram: PF02708; rOmpA\_rOmpB; 1.  
 CC Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;  
 CC Complete proteome.  
 KW SIGNAL 1 38  
 FT CHAIN 39 2021  
 FT POTENTIAL.  
 FT OUTER MEMBRANE PROTEIN A.  
 FT THR-RICH.  
 FT THR-RICH.  
 FT N -> NN (IN STRAIN INDIAN TICK TYPHUS).  
 FT R -> H (IN STRAIN INDIAN TICK TYPHUS).  
 FT MISSING (IN STRAIN M1).  
 FT MISSING (IN STRAIN MOROCCAN).  
 FT VT -> II (IN STRAINS INDIAN TICK TYPHUS).  
 FT D -> A (IN STRAINS INDIAN TICK TYPHUS, M1  
 FT AND MOROCCAN).  
 FT N -> H (IN STRAIN MOROCCAN).  
 FT M -> I (IN STRAIN INDIAN TICK TYPHUS).  
 FT Q -> K (IN REF. 1).  
 FT I -> V (IN REF. 1).  
 FT V -> I (IN REF. 1).  
 FT T -> N (IN REF. 1).  
 FT G -> D (IN REF. 1).  
 FT IS -> VN (IN REF. 1).  
 FT KATLGALIKATTK -> LLQVGGVVKANTIN (IN  
 FT REF. 1).  
 FT N -> D (IN REF. 1).  
 FT V -> I (IN REF. 1).  
 FT N -> D (IN REF. 1).  
 FT VN -> IS (IN REF. 1).  
 FT LLRVGGVVKANTIN -> KATLGALIKATTK (IN  
 FT REF. 1).  
 FT D -> Y (IN REF. 1).  
 FT P -> N (IN REF. 1).  
 FT N -> K (IN REF. 1).  
 FT L -> S (IN REF. 1).  
 FT Y -> S (IN REF. 1).  
 FT K -> Q (IN REF. 1).  
 FT N -> Y (IN REF. 4).  
 FT 1314







Db 1439 NKIRLMGLOVWAGPVSNTGNLKVREGVTVTAASFDNETGAEVNAKSAATLTT-SGAARN 1497  
 QY 1266 STKGQVDLLAONGSAGSINAANVTNTTGLTTVAGSDIRKATS-----GTL 1312  
 Db 1498 AGKMOV-----KEAATITVAASV--SNPGTFT--AGKDIITVTSRGGFNDCKMESNKDI 1546  
 QY 1313 VINAK-----DAK--LNGDASGDSTE-----VNAVNASGSGSVTAATSSV 1351  
 Db 1547 VIKTEQFSNGRVLDAKAKHDLTVAAGQADNRGSLKAGHDFTVQAQRIDNSGTMAGHDAITL 1606  
 QY 1352 NI-----TGDLTVNGLNIISK-----DGRN----- 1372  
 Db 1607 KAPHLRNTGOVAVGHDIHIINSAKLENTGRVDARNDIALDVADEFNTGSLYAEHDATLTL 1666  
 QY 1373 -----TVRLRGKEI--EVKYIOPG--VASVEEVIEAKRVLEKVK 1407  
 Db 1667 AOGTORDLVDDHILPVAEGTLRVKAKSLTTEIETGNPGSLIAEVOENIDNKQAIIVVGK 1726  
 QY 1408 DLS-----DEBRETAKLGVSAREVPEPNTTIVTNTQNEFTTRESSQVILISEGKACFSS 1461  
 Db 1727 DFLSSAHGNVANEANALLWAA-----GELTVKQAN-----ITNKRAALTEA 1768  
 QY 1462 GNGARCTNVA 1472  
 Db 1769 GGNARLTAAVA 1779

RESULT 9  
 HXAL\_HAEIN STANDARD; PRT: 905 AA.  
 AC P44602;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization  
 protein A).  
 GN HXUA OR H10264.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RP [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-Rd / K20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Cotton M.D.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Sutton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,  
 RA Gehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RL Science 269:496-512(1995).  
 CC CC FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.  
 CC CC SUBCELLULAR LOCATION: Secreted.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL: U30712; AAC21929.1; -  
 DR TIGR: H10264; -  
 KW Transport; Signal; Repeat; Complete proteome.  
 FT SIGNAL 1 21 BY SIMILARITY.  
 FT CHAIN 22 905 HEME/HEMOPEXIN-BINDING PROTEIN.

SQ SEQUENCE 905 AA; 98833 MW; 2424013EB437A99D CRC64;  
 Query Match 6.3%; Score 466; DB 1; Length 905;  
 Best Local Similarity 23.7%; Pred. No. 2.5e-12;  
 Matches 251; Conservative 153; Mismatches 347; Indels 308; Gaps 54;  
 QY 72 QCMVYVHGATMQVDGK-TIIRNSVDALIIINWKFNDIQNEMVQFLOENNSAVENRVT 130  
 Db 28 QGSSVVGEARNTVIGNMTIDQKPTQDMHSFDIQNKEVEFPKOPDANSVAYNRVT 87  
 QY 131 NQISQLKGIILDSQVFLINPNTGKDALINTNGTASTLDI-----SNEIKARNFT 186  
 Db 88 GNASQIQGKLTANGKVLANPVGVIITOGAEINVAGLFATTKDLERISENGNGNKFT- 146  
 QY 187 EOTKDKAL---AEIVNHGLITVKGSGVNLIGGVKKEGVI-SVNGSISLLAQCK--IT 240  
 Db 147 RKLKDGQVVRKGVINKEGKIK-AKD-FVVLNGDKVINEGEIDATNGKVYLSGNTFTT 204  
 QY 241 ISDIINPTITYSIAAPENAEVNLGDIFAKGGINVRAATIRNOGKLSADSVSKDSKSNV 300  
 Db 205 LSD-----SSISVALEDNAVQSI-----VQNEGII-----KAGDIT 235  
 QY 301 LSAKEGEAEI-----GGVISAQNOQAKGKIMITGDKVTLLTKTGAVIDLSGKEGTYLG 354  
 Db 236 LNAKGRNQALDSLVMNNGVLEATKVSNGKGVLSADDDVQLNKS--DIKGE--SEVVF- 290  
 QY 355 GDERGEGKNGIOLAKKTSLEKGS-TINVSKEGKGGFAIWMGDIALIDGNINAOGSDIAK 413  
 Db 291 ---TNEPKNKIKITSQTSKVTSPKINFTGKSV-----NIN---GDFGR 328  
 QY 414 TGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSIINAEDPLENNTGINDFPTGTGEASDP 473  
 Db 329 D-----DSKAH--YNEHKKRLDE-----VVIDVPDENIRIAEKDNTGTGTGD- 371  
 QY 474 KNSLKLKTLNTTISNLYKNWNTNITASRLKTVNSNINI-----GSNHLILHSKGQSG 529  
 Db 372 -----SFIQTGALSSLLANNKYNL-KGQDVNISGRHIDSFRGSDSLKLNTQGH- 421  
 QY 530 GGVQID-GDITSKGGNLTYS--GGWVDVHKNTLTDQGFNLNITAAVAF----- 575  
 Db 422 --IKINHADIHSTGRLFFITSLQNEKDSQSDITITDSKINLNGANGWGLRSLDKENCNQ 479  
 QY 576 -----EGGNKARDAANAIVAQ--GTVTITGEGKDFRANVSVNLGTGKGLNIIS--SV 625  
 Db 480 RWCRTETSORKKFVHMNRNVFDDQVDVVVAGGFKKVLNDNV--ATGKTNFYIDGGVSR 537  
 QY 626 NNLTHNLSTINISGNITINOTTRK-----NTSYW-----QTS 658  
 Db 538 NNSRVEY-CVLDLDRKTLSELDOQRRRWKYVNDLDMNKAYWHRFDMFATKNTGRSTI 596  
 QY 659 HDHWNV--SALNLETG-----ANFTFIKYSNSKGLTTOYRSSAGVNF 701  
 Db 597 KQTEINISNKLNGFVHLLAEKIKLDSKIDITFDK---DNSQDISQ-----INR 647  
 QY 702 NGVNGKNSFNLREGAKVNEFKL---PNNMNTSKPLPIRFLANITATGGSVFFDIYAN 757  
 Db 648 LGNNGKYSM-----VNSHIKIVGDEKSDISAKAPYATMTFLIGELIGESKSF--VKS 697  
 QY 758 HSGRGAEKMSLNISNGANFTLNHVRGDDAFKINKOLTINATNSFSLROTCKDDFYDG 817  
 Db 698 HQSY-----TFRTDGTAKIAGKSKDKLKIT----- 723  
 QY 818 YARNATNSVTNISILGNVTLGONSSSITNITEKAANYTLEANNP-----NOONIR 873  
 Db 724 ----AINT-----GGR-----TGKEVI-----INGAFSGIDNDANIA 751  
 QY 874 DRVIKLSLVNSSLTGENADIKONLTISESATPKGTRDPLNITNFTNGTAEINI 933  
 Db 752 NMAFTIGD---NANTKTTEIENAD-----ITALAPNGGATYLS- 785  
 QY 934 TQGV-VKLGNVTNDGDLNITTHAKRNQORSTIGGDI-INKKGSINITDSNNDAEIQIGNI 991  
 Db 786 SKQVEIEVNSNFTFELPREKNFNQTKIKGDSSTKLSRGERFARLYDKINGVR---ASNL 842



```

QY 840 GQNSSITGNTIEKAAVNTLEAN-----NAPNOONIRDVIKIGSLLVN-GSLSLTGEN 894
DB 684 GQVDMLLITGD---DKGKVIHAGHTVFNAGNTYS-----GKTLVNDGLLTIASHT 732
QY 895 ADIKGMLTTSSEATFKGKTRDL---NITGNFT-----NNGTAEINITQ----- 935
DB 733 ACGVTGMSSEVITIANPGTLDILASTNSAGDYTLNALKGDCMRVQLSSDKMFGFTHA 792
QY 936 -----GVVKLGNTVDGDLNTT---THAKRNORSIIIGDIIKKGSLNITDSNDAEIQ 986
DB 793 TGTEFAGVAQLKDSFTFLERDNTAALTHAMLQ-----SDSENTTSVK 834
QY 987 IG-----GNISQKBNLTISDKINITKQITIKKIGDGEDSSDAYTSNANLTIKTKELKL 1042
DB 835 VGSQSIGGLAMNGGTIIFTD---IPAATLAEGYISVDTLVVAGDYTWKGRNYOVNGT 890
QY 1043 EDLSISGFNKAEITAKDGRD-----LITGNSNDGNSGAE---AKTV-----TFN 1083
DB 891 GDVLID-----VPRPWNDPMANPLTLNLEHDDSHGVQLVKAOTVIGSGSLTLR 943
QY 1084 NVKDSKISADGHNVTLSNKKVKTSSSGNGRESNDNDTGLTITAKN-VEVKNKDIITSKTVN 1142
DB 944 DLQDDEVEAD-----KTLHIAQNTGVVAGDYGFRLTTPAGNGLYVN---YGLKALN 992
QY 1143 ITASEKVTITAGSTINATNCKASITTKTGDISGTISGNTVSVSATVDLTKGSKIEAKS 1202
DB 993 IHGGOKLTL-----AEHGGA--YGATADMSAKIGEG-----DLAINTVROVSLSN 1036
QY 1203 GEANVTSATCTIGTISGNTVNTANAGDLTVGNGA--EINATEGAATLTATGNTLTTEA 1260
DB 1037 GONDYQCATVQMGTLRTDADGALGNTRELNSNAIVDLNGS-----TQVETITGQM 1090
QY 1261 GSSITSTKGQVDLLAONGSTAGSINAANVTNTTGLTIVVAGSDIKATSGTLVINAKDAK 1320
DB 1091 GSTVLVFEK-----ALTVNKGGI---SQGLT---GGGNLNVGTGTLAIEGLNAR 1134
QY 1321 LNGDAS--GDSTEVNANVAGSGSVTAATSSSV---NITGDL-NTVNGLNIIISKDGRNVT 1375
DB 1135 YNALTSISPAEVSLENDTOGLGRGNIANDGLLLTKNVTGELRNSISGKIVSATATDVE 1194
QY 1376 LRG 1378
DB 1195 LDG 1197

RESULT 11
OMP_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OmpB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor."
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;

```

---

```

RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: X16353; CAA34403.1; -
DB InterPro: IPR003858; rOmpB_rOmpB.
DB Pfam: PF02708; rOmpA_rOmpB; 1.
DB Antigen: S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1333
FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
FT SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;
QY 34 SEKPARMKVRHLAKPLSAMLISLG-----VTSIPQSVLASGLQGMVHVGTATMOVDG-- 87
DB 2 ACKPNFLK-----KLISAGLVTAATATIVASFAGSAMGAATQONRTTNGAAT-TVDGAG 54
QY 88 -NKTIIIRNSVDALINWKQFNIDQNMVQFLOENNNNSAVFNRTVTSQISOLKILDSNGOV 146
DB 55 FDOTAAPANGVALN-AVITANANGINF---NTPAGSFNGLLLNTANNL----- 100
QY 147 FLINPNTIGTKDAIINTNGFTASTLDISNENIKARNFTFEOTKDKALAEIVNHGLITVG 206
DB 101 -----AVTSED---TTLGF-----INVVHNSHFNLTNAGKTL---TIT 136
QY 207 KGSVNLGGKVKNEGVISV---NG-----GSISLLAGOKITISDIINPTITS 252
DB 137 GCGVTNAQAATAKNAONVVQFNNGAAIDNDLKGVRIDFGAPASTLVFNLANPT--- 192
QY 253 IAAPENEAVNLGD--IPAKG--GNINVRATTRNOGKLSADSYSKDKSONIVLSAKEGA 308
DB 193 ---TQKAPLITGLDNVAVNGVNTLVNNGFI---QVSNKSFATVVKAINI-----A 237
QY 309 EIGGVI-----SAQNOQAKGGKLMITGDKVTLTKTGAVIDLSGKEGETY-----LGG 355
DB 238 DGQGIIFNTDANNANTLNLAQAGTTINFTG---TDGTGRVLVLSKHAHAATNFTITGSLGG 294
QY 356 DERGECK-NGIOLAKKTSLEKSGSTINVSKEKG-----GFAIVW--GDIALIDGNI----- 403
DB 295 NLKGVIEFNTVAVDGQLTANAGANAIVGTNNGAGRAAGFVSVVDNGKVTATIDGOVYAKD 354
QY 404 -----NAQSGD---IAKTGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLF 454
DB 355 MVITQANATQGVNFRHLVDVGADGTTAFKTAASKVITITQDSNFGNTDFGNLAAIKVP-- 412
QY 455 NNTGINDPEPTGTEASDPKKNSELKTLTNTTISVNLKNAWMTNTASRKLTVNSINI 514
DB 413 NATLTGNTF---TGDASNPCNTAGVITFDANGTLESASADA---NVAVTNNITAEASGA 466
QY 515 G-----SNSHLILHSGKQRG-----GKVVQVLSGTHAAELRLGNAGSIFKLADGTVINGKVNQTLVGGALAACTITLDGSATITGD 526
DB 467

```



FT REPEAT 244 252 15 (INCOMPLETE).  
 FT CONFLICT 207 207 D -> DRGDDVTPDD (IN REF. 1).  
 FT CONFLICT 514 514 A -> R (IN REF. 3).  
 FT CONFLICT 1698 1698 D -> N (IN REF. 1).  
 FT CONFLICT 1795 1798 QYLE -> ITLQ (IN REF. 1).  
 FT CONFLICT 1836 1837 SA -> T (IN REF. 1).  
 SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 5.8%; Score 429.5; DB 1; Length 1953;  
 Best Local Similarity 20.4%; Pred. No. 2e-10;  
 Matches 336; Conservative 225; Mismatches 581; Indels 503; Gaps 79;

QY 73 GMDVHGTATQVQDKTKIIRNSVDAILNKKOFNIDONEMVQFLOENNNNSAVFNRTSQ 132  
 DB 310 GWDV-----TDANAALVETQENGLYWKY-----DSRGYLLIADDDNTVVISG--DDQ 354  
 QY 133 ISQLKGLDSGOVFLINPENGITIGKDAINTNGFTASTLD-----ISNENIKARNFTF 186  
 DB 355 AHNSDRGMDISGQ---DRTGVIIISGDRVNTLTGSSVTDGATGMVISGDC--TNTTISG 409  
 QY 187 EOTKDKALAEIVNHGLITVKGDSVNLIGG-----KVNKEGVISVNGSGSISLLA 235  
 DB 410 HSTVDNATGALISNGTNTNFAGDIAVSGGGTAIIIDGNATIKNTGTSIDISG-----A 463  
 QY 236 GOKITISDIINPTTYSIAAPENAEVNLGDFPAKGNINVRATIRNOGKLSADSVSKDK 295  
 DB 464 GSTGTVIDGNARVN-----NDGDMTITDGGT-----GHITGDNVVIDN 503  
 QY 296 SGNIVLSAKEG-----EAEIGGVISAQNOQAKGKL--MITGDKV--TLKGTGAVIDLSGKE 347  
 DB 504 AGSTVSGADATALYIEGDNALVINNEGNOTISGAVGTGTRIDGDDAHTNTGD--IAVDGAG 562  
 QY 348 GGETYLGDERGEKNGIQIAKRTSLFKGSTINVSKEGGEFAIVMGDIALIDG--NINA 405  
 DB 563 SAAVIINGD-----NG-----SLTQA-----GDLVTDGANGIIT 592  
 QY 406 QGSGDIKAGTGFVETSGHDLFIKDNAIVDAKEMLLDFDNYS--INAEPLFNNTGINDEF 463  
 DB 593 YGTCEAKNTG-----NATV-----RDADSVGVVAGEKNTFKNKGDIDVS 633  
 QY 464 PTGTEASDPKKNSELKTTLTNTTISYLNKNAWMTNITASRKLTVNSINIGNSHLIH 523  
 DB 634 LMGTG-----ALVSGDMS-----QVTLGDGINVVS-----VQD 661  
 QY 524 SKG--QRCGGVOIDGITSKGNLTISYSGWVDV--HKNITLDGFLNITAAVAPE-- 576  
 DB 662 SEGVSSTYGVSVGDSNA-----VDITGNVNISADYGQDDLAAGAPPLTGVV 709  
 QY 577 -GGNNKARDAANAKIVAOGVTITIGEGKDFRANNVSLNGTKGLNLTISVNNLTHN---L 632  
 DB 710 VGGNGNTVTLNGLNIDNDLSATG--GOYLDVGLSVTGDNDNVEIDGGI--NITHSEDPL 767  
 QY 633 SGT-----INISGN--ITINQTRKNTSYWQTSH-----DSHWNVYAL 668  
 DB 768 DQTSADITGISVSGNSTVTLNGHSTIDTNTVVGHVILARVNNVGGSLILGDDSVVDNVVS 827  
 QY 669 NLETGANFTFKYISSNSKGLTTOYRSSAGVNFNGV-----NGNMSFN----- 711  
 DB 828 YIPTG--YTYTNALLMADGEG--TSIENKGDITSHGVSIVIRADNGSEVNSGDIILVYATS 884  
 QY 712 -----LKEGAKVNFKLPKNENMTSKPLPIRFLANTATATGGGVS----- 750  
 DB 885 SNSSEDRAAITRASGEVAVHANKAGD-----TTLISDQTPGGSGGIEVYPLKWY 934  
 QY 751 ---FFDIYVNHSGRAELKMSINISNGANFTLNHVRGDDAFKI--NKDITINATNSNF 805  
 DB 935 THTFYAMASDYG-----DVVNEGATI--HLQAGAGYGVYTSASRGALNEGNYL 982  
 QY 806 -SLRQTKDDFYDCYARNAINST-----YNISILGNNVTLGGONSSSITGNTIE 854  
 DB 983 DGLVPLTDD-----ENNTITSYQWPSLYLTSAGMVGAGTDAG--DATAINTGNITV- 1034  
 QY 855 KAANTVLEANNAPNOQNTDRVILKUGSLVNGSLTGTENADIKNLNLTISESAFEGKTR 914

DB 1035 ----NNAG-----FGMALNGGTALINGVITLTAD--DGVTG 1065  
 QY 915 DTLNITGNFTNNGTAETNITQGVVKL-----GNVTNDGDLNITTHAKRNORS 961  
 DB 1066 QADELVGMAALNGGVINISGVINIDADYGOAFSLSSSYIINNGSINLNGSPMDTDS 1125  
 QY 962 IIGGDIINK-----KGSNLNTDSNDAEIOIGNISOKEGNLTISS----- 1003  
 DB 1126 HMGTPPTDKIWIQSLPGSGSDTSTDTFTAGTTLA--NYGTETLNGDVVDVNGWLNEA 1184  
 QY 1004 --KINITKQITIKKIGDGEDS-----SSDATSNANLTIKTKELKTED--LSISGNKAEIT 1056  
 DB 1185 GASLTVNGTVTINGANALANYGLDADAISTWHSLSFNEADGSITDILLUNG----- 1237  
 QY 1057 AKDGRDLTIGNSND--GNSGAEAKTVTFFNNYKDSKISADGHNVTLNSKVKTSSNGGRES 1114  
 DB 1238 ----DVTFFYNGDFTGSIAGTSYQOEIVNTGDMTVAEDGKSLV-----SGSFYF 1282  
 QY 1115 NSDNDTGTITAKNVENKRDITSLKTVNIT--ASEKVITTTAGSTINATNGKASITKTGD- 1172  
 DB 1283 YNEEDATLTNSGSAVEGGEN---TIINLRANDSLTQVNSGTITATNGYSAITVNGSN 1338  
 QY 1173 -----ISGTISG-----NTVSVA-----TVDLTT 1192  
 DB 1339 DPKWIWNTATGVINGINPDAPLINLGRYNGFNGOCTINVOGDNAVAISGGTSSVYINLVN 1398  
 QY 1193 KSGSKIEAKSGEANTVSATGTIGTISGN--TVNVYTANA-----GDLT--VGN 1236  
 DB 1399 SGTINVTGEOGEDGTNGTGLIGIKGNGNATINNTADGVINVYADDSYAFGKTKAIIN 1458  
 QY 1237 GAEIN--ATEGAATLTATGNTLTTEAGSSITSTKGOVDLLAONGSIA--GSINAAVNTLN 1292  
 DB 1459 NGEINLLCDSGCDITAPGTTGTQ-----NDHNGTADIVIPATTAPEGSITPPADPN 1512  
 QY 1293 TTGTLTT--VAGSDIKATSGTLVINAKDLKNGDASGDSTEVNNAVNASGGSVT-----AA 1346  
 DB 1513 APOQLSNYITGVNADGSSGTL-----KANLVIGDNVYKVDIGFTSGTADTVVVDNAF 1565  
 QY 1347 TSSSV-----NITGDLTVNGLNIIISKGRNTVRLRKETEVEVYIQPGVASVEVEIAKRV 1402  
 DB 1566 TGSNIOGADNITSTSVYVNAQSQDAGHVDVTM-----TRNAVADVATDSVSDVAQA 1619  
 QY 1403 LEKVKDLSDDEERETLAKLV-----SAVRFEVPNNITVTNQ-----NEFTTRPSQVII 1452  
 DB 1620 LD--AGYTNNELYSLVNGTTAELNSALKOVSGAQATTVFREARVLSNRFTMLADAAPOI 1677  
 QY 1453 SEGKACFFSSNGARVCTNVADGQP 1477  
 DB 1678 KDGLA-----FNVYAKGDP 1691

## RESULT 13

OMP\_B\_RICCN STANDARD; PRT; 1655 AA.  
 ID OMPB\_RICCN 09KX43; 09KX45;  
 AC 09KX43; 09KX45; 09KX45;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (surface protein antigen) (Cell surface antigen 5) (Sca5) (rompB)  
 DE antigen) (120 kDa outer membrane protein) (Surface protein  
 DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).  
 GN OMPB OR RC1085.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OC NCBI\_TaxID=781;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,





Db 1045 RFVDG-ILSHSTSMITTKANNQTVYLGNAFVGNIGDSD-----TPVASVFTGSD 1095  
QY 1110 GGRESNDNDTGLTITAKNVEVNDKITSKVTNITASEKVTITAGSTINATNGKASITTK 1169  
Db 1096 GG-----AGL-----CGNIYSQVIDEFTYNLGISNSVNLGGGT-TAINGKINLRN 1141  
QY 1170 TGDIT-SGTIS-GNTVSASFVDLTKS-GSKIEAKSGEANVTISATGCTIGSTGNTVNT 1226  
Db 1142 TLFASGTSTWGNSTSTETTLTLANGNIGNIVILEGAQVNAIT-TGT-----TTIKVQ 1193  
QY 1227 ANA-----GDLT---VNGAEINATGAATLTATGNTLTTEAGSSITSTKGOVDLLAQNG 1278  
Db 1194 DNANFSGTQTVTLIQGGARFNGTLGGPNFVVT-----GSRNFVNYGLIRANODY 1245  
QY 1279 SIAGSINAANVTNTGTLTAVAGSDIKATSG-----TLVINAKDAKNGD---ASGDST 1330  
Db 1246 VIPTNNAENVIN-----DIANSFGGAPGVQNVTTFVNATNTAAYNNLLAKNSAN 1299  
QY 1331 EVNAVASGSGSVTAATSSVNTIGDL-----NTVNGCLNIISKDGRNTVRLRGKETEYKV 1385  
Db 1300 SANFVGAIVTDSNAITNAQDLVAKDIOAQLGNRLGALRYLGT---ETAEMAGPEAGA-- 1355  
QY 1386 IQPGVASVEEVIE 1398  
Db 1356 IPAAVAAGDEAVD 1368

RESULT 14  
HLVA\_SERMA STANDARD; PRT; 1608 AA.  
AC P15320;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Hemolysin precursor.  
GN SHLA  
OS Serratia marcescens.  
SC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
CC Serratia  
OX NCBI\_Taxid-615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
RC STRAIN=SN8; 17037; PubMed=3290200;  
RA MEDLINE=86257037; Braun V.;  
RA Poole K., Schliebel E., Braun V.;  
RT "Molecular characterization of the hemolysin determinant of Serratia  
marcescens."  
RL J. Bacteriol. 170:3177-3188(1988).  
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
DEFINED.  
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA  
REQUIRES SHLB FUNCTION.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M22618; AAA50323.1; -  
DR PIR; A28182;  
KW Hemolysin; Toxin; Outer membrane; Signal.  
FT SIGNAL 1 30  
FT CHAIN 31 1608 HEMOLYSIN.  
SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 5.7%: Score 425; DB 1; Length 1608;  
Best Local Similarity 21.0%; Pred. No. 2.4e-10;  
Matches 351; Conservative 216; Mismatches 576; Indels 528; Gaps 79;  
QY 2 NKIYPLKSKRLNALVAVSELARCDHSTEGSEKPKRMKVRHLALPLSLAMLLSLGVT 61  
Db 3 NNPFSLSAAGKLAALAILAASAGAYAE-----IVAANGANG 41  
QY 62 IPQSVLASGLQMDVHCTATWQVDGNTKIIRNSVDAILNKQFNIDQ----- 109  
Db 42 PGVSTAATAAGVVDIV-----APNGN-GLSHN-----QYQDFNVNPGAVLNSREAG 88  
QY 110 -NEMVFOLOENNN-----SAVFNRTSNQISQLKG-----ILDSGOVFLIIPNGITGKD 159  
Db 89 LSQLAGQLGANPLNGREASVILNEVIGRNPSSLHGQOEIFGMAADYVLANPNGISCOSC 148  
QY 160 AIINT-----NGFTASTLDISNENIKARNFT----- 185  
Db 149 GFINTSHSVVGNPLVNGVLYQGYSTFCNRNTLSLNGTLNAGGVLDLIAPIKIDSRGEVI 208  
QY 186 ---FQTKDKALAEIVN---HGLITVKGDSVNLIGGKVKNEGVISVNGSGISLLAGOKIT 240  
Db 209 VQDFKQSGKVTSAAINAISGLNRVARDGT-----QASQQMPTALDSYVLGSMQAGRIN 263  
QY 241 ISDIINPTITYSIAAPENEAENVLDIFAKGGNINVRATIRNOGKLSADSVSKDSGNIV 300  
Db 264 I-----INTAOGSGVKLAGSLNAGDELKVKAYDIRSESRV--DDASSNKGMDN 310  
QY 301 LSAKEGEAEIGVISNAQ---NOQAKGGKLM-----ITGDKVTLTKTGAVIDLS 344  
Db 311 YONRGYVNDRSSQTLTRTELKGNISLVADNHAHTATDIRGEDITLQ-GGKFLTD 369  
QY 345 GKEGGTYLGGDERG---EGKNGIQIAKTSLEKSTINVSIGKEGFAIWGDIALIDGN 402  
Db 370 GQOLKOTGHTDDRFYFSDYDVTREORELOQAGSTVAASGSAK---LISTQEDVKLLGAN 427  
QY 403 INAOGSGDIAKTGGFVETSGHDLFIKDNAIVDAKEWLLDFDNVSNIAEDPLFNNTGIND 462  
Db 428 VSADRALS-KAARDVHLAG--LVEKD----- 451  
QY 463 FPTGTGEASDPKKNSELKTTLTNTTISNYLKNATWMTNITASKRLVNSNINIG--SNSHL 520  
Db 452 -----KSSE-----RGYQRNH-----TSSLRTGRWNSD- 475  
QY 521 ILHSGORGGVQVQIDGDTISKGNLTYSGGWYDVHKNITLD-----QGFLNITAAVAF 575  
Db 476 --ESELKASELSRSEGLTLKAGRNVTQCAKVHAQRDLTIDADNOIQVGVQKATANAKAV 533  
QY 576 E-----GNNK-----ARDAANA-KIVAQGT-----VTITGEGKDFRANNVSL 612  
Db 534 RDDKTSWGGIGGDNKNNSNRREISHASELTSGTLRLNGOQGVITG-SKARGQKGEV 592  
QY 613 NGTGKGLNIISNNLTHNLSTGNTINOTTRKNTSYWQTSKSHSNVSAINLET 672  
Db 593 TATHGGURIDNALSTTVDKIDARTGAFNIT--SSSHKADNSYQSS-----TASLKS 643  
QY 673 GANFTFTKYISSNSKGLTTQYRSSAGVFNMGVNGMNSFNKEGAKVNFKLPKNENMTSK 732  
Db 644 DTNLTLSHKDADVIG--SOVASGGELSVESKGTGNI--NVKAAER-----QONIDEQK 692  
QY 733 PLPIRFLANITATGGGVFFDIYANHSQ---RGAELKMEINISNCAFTLNHVR---G 786  
Db 693 -----TALTUNG-----YAKEAGDKQYRAGLRIEHTROSEKTRTTRTENSASSLSGG 737  
QY 787 DDAFKINKDLTIN-----ATNNSFSLRQTKDDFYDGVYARNAINSTYISILGGNVTLGGQ 841  
Db 738 SVKLKAEKDVTFSSKLVADKGDASVSGNVSFLAADDKATASN-TEOTKIGGGFYITGG- 795  
QY 842 NSSSITGNTIETIEKAANVTLEA---NNAPNOONIRDRVIKGLSLVNGSLTGENADIK 898  
Db 796 -----IDKLGS--GVEAGYENKTKQAQS--SKAITSQS-----DVK 827  
QY 899 GNLTISESATPKGKTRDTLNTITGNTNNGTAEINITQGVVGLGNVTNDGDLNITTHAKRN 958



```

Db 54 --GPDQAAGANIPVAPNSVI--TANANNPTFN--TPN-----GHLSN--LFLDTAND 98
Qy 153 -GITIGKDALINNGFTASTLDSINENIKARNFTFOTRKALAEIVNGLITVGRGVS 211
Db 99 LAVTINED---TTLGF-----JNTIAOQAFNFVAAKIL-NITGOGITVOEASNTI 148
Qy 212 NLIGGKYNKGVISVNG-----GSLLAGQKITSDIINPTITVSIARAPEANVLGD 265
Db 149 NAQNALTKVHGGAANAINLSGLSITFAAAPSVLLENLNP--TOEAPLTIGANSKI 205
Qy 266 IFARGGNINRAATIR-----NOGKISADSUSK-----293
Db 206 VNGNGTLLNTINGFIOVSDNTEAGIKTINDDCCGLMFNSTPDAANTLNLQVGGNTINFN 265
Qy 294 --DKSGNIVLSAKBGEA-----EIGGVISAQNOQAKGKLMITGDK-----332
Db 266 GIDGTGLVLVSKNGAATEFNVTGLGNLKGIIEL-NTAAVAGKLIISOGGAANAVIGTD 324
Qy 333 -----VTLKTGAVIDLSGKEGETYL-----GGDERGEGKNGIOLAKKTSLEKG 376
Db 325 NGAGRAAGFIVSDNGNAATISGOVYAKNMVIOANAGQGVTEHIVDVGLGCTINF-KT 383
Qy 377 STINVSKEKGEA-----IVMGDIALIDGNINAOGSGDIAKTGGFVETSGHDLF 426
Db 384 ADSKVIITENSNGSTNFGNLDTOIVVPDKILKGNF-----IGDVKNNG--NTAGVITF 436
Qy 427 IKDNIYDAKEWLLDFDNVSNINAEDEPLENTGINDPEFTGTGEASDPKKNSELKTLTNT 486
Db 437 NANGALVSA-----STDPIAVTNINAIABEGGVVE-----LSGI 472
Qy 487 TISNY-LKNAMTNITASKRITVNSINSNGSHLILHSGQORGGVQIDGD--ITSKGG 543
Db 473 HIAELRLGNGSSI-----FKLADGTVINGPVNQALMNNALAAAGSIOLOGSAITIGD 527
Qy 544 NLTIYSGGVVDVHKNTLIDOGFLNITAASVAFEGGNKARDAANAKIVAQ--CTVTITG 600
Db 528 N-----GGVNAALQHITL-----ANDASKILALDGANIIGANVYGA-IFHQANGGTIKLTN 577
Qy 601 EGKDFRAN---NVSLNGTGKGLNISSVNNLTHLSGFI-----NI-----638
Db 578 TONNIVVNFDLDTITDKTGV-VDASSLTNNQTLTNGSIGTVVANTKTLAQLNIGSSKTI 636
Qy 639 --SGNITINOTRKNKTSYQWTSQSHSNVNSALNLETGANFTFIKYSNSKG-----688
Db 637 LNAQDVAINELVIENNGSVQLNHN-----YLITKTINAANOQIIVAADP 682
Qy 689 LFTQYRSSAGVNFVNGVNMSPFLKEGAKVNFKLKPNENMNTSKPLPIRFLAN---ITAT 745
Db 683 LNTNTLADGTLNLSAENPLS-----TIHFATK-----AANADSLNV 720
Qy 746 GGSVFFDIYAHNSGRGAELKMSSEINSGANFTLNSHVGDGDAFKINKDLTINATNSF 805
Db 721 GKG---VNLIANITN-DANVGSJHPSGGTSIVSGTVGGOOGHKLANLILDNGTIVKF 776
Qy 806 -----SLROTKDDFDYDGYARNAINSTYINISILGN---VTLGGQNSSSI 847
Db 777 LGDTFTNGGKTEGKSILQISNNYTTDHSVADN-TGTLEFVNTDPTITVLNKGAYEGV 835
Qy 848 -----TGNITIEKAANVTLEANNAPNOQNIIRDVILKGLSLVNGSLSTGENADIKG 899
Db 836 LKQVIISGPGNIVFNEIGNVIVHGAAN-----SISFENASLGT 875
Qy 900 NLTISESATFKGTRDTLNI---TGNFT--NNGTABINITQGV---VKLGNVTND-----946
Db 876 SLFLP-----SGTPLDLVLTIKSTVGNGTVDFNENAPIVVVGSDSMINNGQIIGKKNITIA 930
Qy 947 ---GDLNITTHAKRNORSIGGDIINKKSLNITDSNDAEIQIGGNISQKEG-----N 997
Db 931 LSLGSDNSITVNAVNTLYSGI-----RTTKNNOGTVTLSSGMPNPNPGTIYGIG 977
Qy 998 LTISDDKINIKQITIKKID--GEDSSSDATSNANLTIKTKELKLTEDLSISGPNKAEI 1055

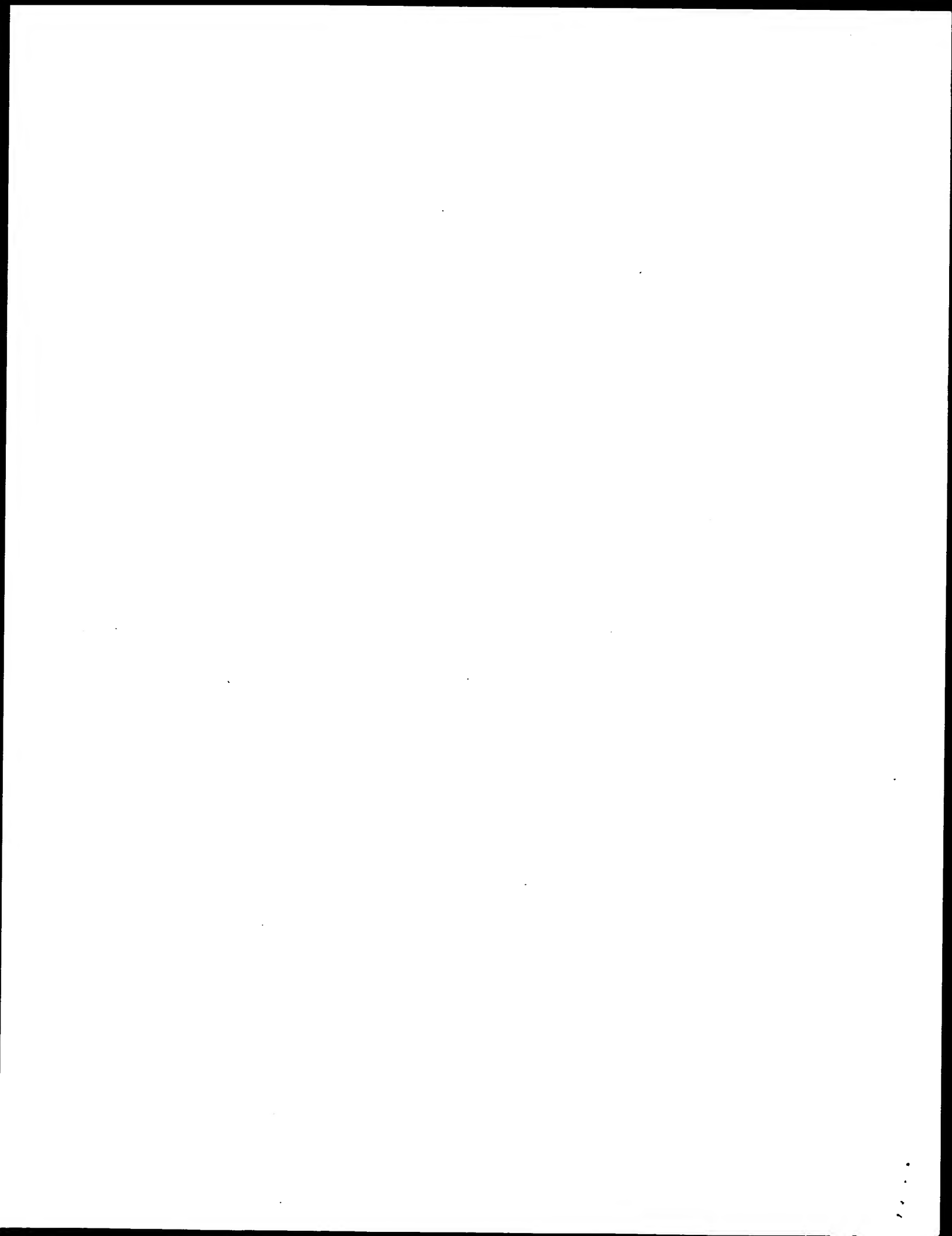
```

```

Db 978 LENGSPKL---KQVFTTDDYNLGSIIANNVTINDYVTLTT-----GGIAGTDF 1023
Qy 1056 TAKDGRDLTIGNSDNGSGAEAKTVTFENNVDKSKISADGHNVTLNSKVKTSSSNGG---- 1111
Db 1024 DAK---ITLGSVN-GNANVREVDSTFSDPRSMIVATOANKGTVTYLGNALVSNIGSLDT 1078
Qy 1112 ----RESNDNDTGL--TITAKNVEVNKDIITSLKTVNITASEKVTITTAGSTINATNGKA 1164
Db 1079 PVASVREFTGNDSGAGLOGNIYSONIDFG-----TYNLTILNSNVILGGGT-TAINGEI 1130
Qy 1165 SITTKTGDISGTIS--GNITVSVSATVDLTTKS-GSKIEAKSGEANTVSATGTIGTISGN 1221
Db 1131 DLTNNLIFANGTSTWGDNTSISTTLNVSSGNIGQVVAEDAQVNAATT-TGTTTIKIQDN 1189
Qy 1222 -TVNVTANAGDLTVNGAEINATEGANLTATG-----N 1254
Db 1190 ANANFSGTOAYTLIOGGAREFNGTLGAPNFAVTGSNIFVKYELIRDSNQDYVLTRTNDVLN 1249
Qy 1255 TLTTEAGSSITSTKGOVD-----LLAQNGSIAGSINAANVTTLNTTGT 1296
Db 1250 VVTAVGNSAIAANAPGVSONISRCLESTNTAAYNNMLLAKDPDSVATFVGA-IATDTSAA 1308
Qy 1297 LTTVAGSDIKATS-----GTL--VINAKDAKLNGDA-----SGDSTEV-----1332
Db 1309 VTTVNLNDTQKTQDILLSNRLGTLRLYLSNAETSDVAGSATCAVSSGDEAEVSYGVNAKPEY 1368
Qy 1333 NAVNASGSGSVTAATSSSVNITGDLNTVNGLNII 1366
Db 1369 NIAEQDKKGAGIYAKKAKTTGVVVVGLDTLASDNLN 1402

```

Search completed: March 24, 2003, 15:22:57  
Job time : 36 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:22:34 ; Search time 47 Seconds  
(without alignments)  
6475.140 Million cell updates/sec

Title: US-10-092-880-4  
Perfect score: 7407  
Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	7318	98.8	1477	2	Q48028 haemophilus
2	4957	66.9	1536	2	Q48031 haemophilus
3	4711	63.6	1557	2	Q9RNI2 haemophilus
4	900	12.2	2055	2	Q93DC7 haemophilus en
5	876	11.8	1910	16	Q82BY3 versinia pe
6	785	10.6	1371	16	Q8XQ42 ralstonia s
7	744.5	10.1	2154	16	Q9HYG6 pseudomonas
8	636.5	8.6	1417	16	Q9HVG6 pseudomonas
9	602	8.1	3705	16	Q8ZHA1 versinia pe
10	594	8.0	3705	2	Q9F285 versinia pe
11	593.5	8.0	4919	2	Q9ZHL0 haemophilus
12	561	7.6	2737	16	Q8XP07 haemophilus
13	553.5	7.5	1018	16	Q9HW06 pseudomonas
14	548	7.4	3930	16	Q9BE20 rhizobium l
15	543	7.3	3165	16	Q8RDQ9 fusobacteri
16	542.5	7.3	2143	16	Q8RIP5 fusobacteri

17	538.5	7.3	1881	16	Q8RGK2	Q8rgk2 fusobacteri
18	531.5	7.2	1270	16	Q8XAN9	Q8xan9 escherichia
19	530	7.2	3501	16	Q8Y106	Q8y106 ralstonia s
20	529	7.1	3552	16	Q8XSD6	Q8xsd6 ralstonia s
21	523.5	7.1	4152	2	Q9ZHL3	Q9zhl3 haemophilus
22	520	7.0	1975	16	Q9K0S7	Q9k0s7 neisseria m
23	520	7.0	2535	16	Q8ZDR6	Q8zdr6 versinia pe
24	517.5	7.0	1995	16	Q9JY23	Q9jy23 neisseria m
25	512.5	6.9	2340	16	Q9ZD91	Q9zd91 rickettsia
26	508	6.9	1604	2	Q9KK99	Q9kk99 rickettsia
27	508	6.9	2015	16	Q9JRD2	Q9jrd2 neisseria m
28	505.5	6.8	2712	16	Q9F3X5	Q9f3x5 pasteurella
29	504	6.8	3295	16	Q8Z1A8	Q8z1a8 versinia pe
30	501.5	6.8	1480	2	Q8VNR0	Q8vmr0 photorhabdu
31	495.5	6.7	3155	16	Q8XYI3	Q8xyi3 ralstonia s
32	491	6.6	3233	2	P71401	P71401 haemophilus
33	489	6.6	3241	2	Q9AHF9	Q9ahf9 fusobacteri
34	485	6.5	2059	16	Q9ED50	Q9ed50 xylella fas
35	482.5	6.5	3322	16	Q8XQZ5	Q8xqz5 ralstonia s
36	480.5	6.5	2026	2	Q9JFJ4	Q9jfd4 neisseria m
37	480.5	6.5	2039	16	Q8ZN57	Q8zns7 salmonella
38	480	6.5	1594	2	O32608	O32608 edwardsiell
39	479	6.5	2462	16	Q8RG23	Q8rg23 fusobacteri
40	478	6.5	2751	16	Q8XUK0	Q8xuk0 anabaena sp
41	477.5	6.4	1487	16	Q8YK40	Q8yk40 anabaena sp
42	475.5	6.4	1749	16	P73032	P73032 synechocyst
43	474	6.4	5291	16	Q8XZT1	Q8xzt1 escherichia
44	472.5	6.4	1152	16	Q8YZU1	Q8yzu1 anabaena sp
45	471	6.4	2514	16	Q9JY30	Q9jy30 neisseria m

## ALIGNMENTS

### RESULT 1

Q48028 PRELIMINARY: PRT: 1477 AA.  
ID Q48028  
AC Q48028:  
DT 01-NOV-1996 (TremBLrel. 01, Created)  
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)  
DE Adhesin.  
GN HMW2A  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus  
OX NCBI\_taxid=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=12; PubMed-1548058;  
RX MEDLINE=92192797; PubMed-1548058;  
RA Barenkamp S.J., Leininger E.;  
RT "Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis.";  
RT Infect. Immun. 60:1302-1313(1992).  
RL EMBL: U08875; AAA20524.1;  
DR InterPro: IPR001000; Glyco\_hydro\_10.  
DR InterPro: IPR001069; SHprot\_acsite.  
DR PROSITE: PS00591; GLYCOSYL\_HYDROL\_F10; UNKNOWN\_1.  
DR PROSITE: PS00639; THYOL\_PROTEASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 1477 AA; 154473 MW; B057C23F1AD24B0E CRC64;

Query Match 98.8%; Score 7318; DB 2; Length 1477;  
Best Local Similarity 99.1%; Pred. No. 1.7e-231;  
Matches 1464; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MNKIYRLKFSKRLNALVAVSELARGCDHTEKSEKPARMKVRLHAKPLSAMLISLQVT 60  
Db 1 MNKIYRLKFSKRLNALVAVSELARGCDHTEKSEKPARMKVRLHAKPLSAMLISLQVT 60  
QY 61 SIQSVLASGLQGMVDVHGTATMQVDGNKTIIRNSVDAINNKQFNIDQNVQFLOENN 120



Db 61 SIPOSVLASGQMDVYVHGATATQVGNKTIIRNSVDALINNKQENIDQENWQFLQENN 120  
QY 121 NSAVFNRTVSNQISQKGLDSDNGQVFLNPNIGITIGDKAIINTNGFTASTLIDISNENIK 180  
Db 121 NSAVFNRTVSNQISQKGLDSDNGQVFLNPNIGITIGDKAIINTNGFTASTLIDISNENIK 180  
QY 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKGVKNEGVISVNGGSIISLLAGQKIT 240  
Db 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKGVKNEGVISVNGGSIISLLAGQKIT 240  
QY 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300  
Db 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300  
QY 301 LSAKEGEAEIGGVIISAQOQAKGGLMITDKVTLTKTGAVIDLSKGEGETYLLGDERGE 360  
Db 301 LSAKEGEAEIGGVIISAQOQAKGGLMITDKVTLTKTGAVIDLSKGEGETYLLGDERGE 360  
QY 361 GKNGIOLAKKTSLEKGSINVSKEKGGFAIVWGDIALIDGNINAGSGDIAKTGGFVET 420  
Db 361 GKNGIOLAKKTSLEKGSINVSKEKGGFAIVWGDIALIDGNINAGSGDIAKTGGFVET 420  
QY 421 SGHDLFIKDNAIVDAKELLDFDNYSINAEPLFNNTGINDFPTGTGEASDPKKNSELK 480  
Db 421 SGHDLFIKDNAIVDAKELLDFDNYSINAEPLFNNTGINDFPTGTGEASDPKKNSELK 480  
QY 481 TLTNTTISNLYKNATWNITASRLKTLVNSNITSGNSHLILHSKGGQVQIDGITS 540  
Db 481 TLTNTTISNLYKNATWNITASRLKTLVNSNITSGNSHLILHSKGGQVQIDGITS 540  
QY 541 KGGNTIYSGWVDVHKNTLDDQGLNITAAVAFEGGNKARDAANAKIVAQGVTVITG 600  
Db 541 KGGNTIYSGWVDVHKNTLDDQGLNITAAVAFEGGNKARDAANAKIVAQGVTVITG 600  
QY 601 EKGDFRANVNSLNGTGGKLNIISSVNNLTHNLSTINISGNITINOTTRKNTSVWQTSHD 660  
Db 601 EKGDFRANVNSLNGTGGKLNIISSVNNLTHNLSTINISGNITINOTTRKNTSVWQTSHD 660  
QY 661 SHWVNSALNLETGANFFIKYISSNSKGLTQYRSSAGVNFNGVNGNMFNLKEGAKVNF 720  
Db 661 SHWVNSALNLETGANFFIKYISSNSKGLTQYRSSAGVNFNGVNGNMFNLKEGAKVNF 720  
QY 721 KLKPNENMNTSKPLPIRLANITATGGSVFFDIYAHNSGRGAEKLMSEINISNGANFTL 780  
Db 721 KLKPNENMNTSKPLPIRLANITATGGSVFFDIYAHNSGRGAEKLMSEINISNGANFTL 780  
QY 781 NSHYRGDDAFKINKDLINATNSNFSLRQTKDDFYDGYARNAINSTYNTSILGNNVTLGG 840  
Db 781 NSHYRGDDAFKINKDLINATNSNFSLRQTKDDFYDGYARNAINSTYNTSILGNNVTLGG 840  
QY 841 QNSSSSITGNITTEKAANVTLEANNAPNOQNIIRDVTKLGSLLVNGSLTGENADIKGN 900  
Db 841 QNSSSSITGNITTEKAANVTLEANNAPNOQNIIRDVTKLGSLLVNGSLTGENADIKGN 900  
QY 901 LTISEATFKGKTRDTLNTGNFTNNGTAEINITQGVVYKLVNVTNDGDLNITTHAKRNOR 960  
Db 901 LTISEATFKGKTRDTLNTGNFTNNGTAEINITQGVVYKLVNVTNDGDLNITTHAKRNOR 960  
QY 961 SIIGGDIINKKGSINLTDSDNDAEIQIGGNIQKGNLISSEKINITQITIKKGIDGE 1020  
Db 961 SIIGGDIINKKGSINLTDSDNDAEIQIGGNIQKGNLISSEKINITQITIKKGIDGE 1020  
QY 1021 DSSSDATSNANLTKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGSGAEAKTV 1080  
Db 1021 DSSSDATSNANLTKTKELKLTEDLSISGFNKAETAKDGRDLTIGNSNDGSGAEAKTV 1080  
QY 1081 TFNNVDSKISADGHVNTLSKVKTSNGSGRESNDNDTGLTITAKNVEVNDITSLKT 1140  
Db 1081 TFNNVDSKISADGHVNTLSKVKTSNGSGRESNDNDTGLTITAKNVEVNDITSLKT 1140  
QY 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATGDLTTKSGSKIEA 1200  
Db 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATGDLTTKSGSKIEA 1200

Db 1141 VNITASEKVTTTAGSTINATNGKASITTKTGDISGTISGNTVSVSATGDLTTKSGSKIEA 1200  
QY 1201 KSGEANTVSATGTIGTIGTISGNTVNTANAGDLTVGNGAEINATGAATLTATGNTLTTEA 1260  
Db 1201 KSGEANTVSATGTIGTIGTISGNTVNTANAGDLTVGNGAEINATGAATLTATGNTLTTEA 1260  
QY 1261 GSSITSTKGGVDDLLAQNQSGTAGSINAANVTINTTGTTLTTVAGSDIKATSGTLVINAKDAK 1320  
Db 1261 GSSITSTKGGVDDLLAQNQSGTAGSINAANVTINTTGTTLTTVAGSDIKATSGTLVINAKDAK 1320  
QY 1321 LNGDASGDSSTEVNANVNASGSGSVTAATSSSVNITGDLNTVNGLNILSKDGRNVTURLGKE 1380  
Db 1321 LNGDASGDSSTEVNANVNASGSGSVTAATSSSVNITGDLNTVNGLNILSKDGRNVTURLGKE 1380  
QY 1381 IEVKYIOPGVASVEEVEAKRVLEKVKKDLSDERETLAKLGVSAVRFEVPEPNNTITVNTON 1440  
Db 1381 IEVKYIOPGVASVEEVEAKRVLEKVKKDLSDERETLAKLGVSAVRFEVPEPNNTITVNTON 1440  
QY 1441 EFTTRPSSQVILIEGKACFSSGNGARVCTNVADGGP 1477  
Db 1441 EFTTRPSSQVILIEGKACFSSGNGARVCTNVADGGP 1477  
RESULT 2  
Q48031 ID Q48031 PRELIMINARY; PRT; 1536 AA.  
AC Q48031; MEDLINE=92192797; PubMed=1548058;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Adhesin.  
GN HMWIA.  
OS Haemophilus influenzae.  
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=12;  
RX MEDLINE=92192797; PubMed=1548058;  
RA Barenkamp S.J., Leininger E.;  
RT "Cloning, expression, and DNA sequence analysis of genes encoding  
RT nontypeable Haemophilus influenzae high-molecular-weight surface-  
RT exposed proteins related to filamentous hemagglutinin of Bordetella  
RT pertussis.";  
RL Infect. Immun. 60:1302-1313(1992).  
DR EMBL: U08876; AAA20527.1; -;  
DR InterPro: IPR000169; ShpO1\_acsite.  
DR PROSITE: PS00639; THIOLEPROTEASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 1536 AA; 159916 MW; 5CA1C31F9DCF188E CRC64;  
Query Match 66.9%; Score 4957; DB 2; Length 1536;  
Best Local Similarity 67.3%; Pred. No. 2.le-154;  
Matches 1046; Conservative 146; Mismatches 265; Indels 98; Gaps 20;  
QY 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLAKPLSAMLISLGV 60  
Db 1 MNKIYRLKFKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLAKPLSAMLISLGV 60  
QY 61 SIPOSVLASGQMDVYVHGATATQVGNKTIIRNSVDALINNKQENIDQENWQFLQENN 120  
Db 61 SIPOSVLASGQMDVYVHGATATQVGNKTIIRNSVDALINNKQENIDQENWQFLQENN 120  
QY 121 NSAVFNRTVSNQISQKGLDSDNGQVFLNPNIGITIGDKAIINTNGFTASTLIDISNENIK 180  
Db 121 NSAVFNRTVSNQISQKGLDSDNGQVFLNPNIGITIGDKAIINTNGFTASTLIDISNENIK 180  
QY 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKGVKNEGVISVNGGSIISLLAGQKIT 240  
Db 181 ARNFTFQTKDALAEIVNHGLITVKGDSVNLIGKGVKNEGVISVNGGSIISLLAGQKIT 240  
QY 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300  
Db 241 ISDIINPTITYSTAAPEAVNLGIDIFAKGGNINVRATIRNOGKLSADSVSKDGSNIV 300

Db 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSXKDSGNIV 300  
Qy 301 LSAKEGAEIIGVYISAOQOAKGKLMITGDKVTLTKTGAVIDLSGEGGETYLGGERGE 360  
Db 301 LSAKEGAEIIGVYISAOQOAKGKLMITGDKVTLTKTGAVIDLSGEGGETYLGGERGE 360  
Qy 361 GKNGIOLAKTSLEKSTINVSKEKGGPAIVMGDIALDGNINAOGSDIAKTGGEVET 420  
Db 361 GKNGIOLAKTSLEKSTINVSKEKGGPAIVMGDIALDGNINAOGSDIAKTGGEVET 420  
Qy 421 SGHDLFTKDNAIVDAKWLDFDNVSNIAEDPLFNWTFINDEPTGTG-EASDPKKNSEL 479  
Db 421 SGHDLFTKDNAIVDAKWLDFDNVSNIAEDPLFNWTFINDEPTGTG-EASDPKKNSEL 479  
Qy 480 KTLTNTTISNYLKNAMWTMTASRKLTVNSSINISNSHLIHSKQORGGVQIDGDI 539  
Db 479 KTLTNTTLESILKKGTFVNITANORIYVNSSINL-SNGSLTLWSEGRSGGVIEINNDIT 537  
Qy 540 ----SKGNLTISYGGVVDVHKNTLD-QGFLNITA-ASVAFEGGNKKAADAANAKIVAQ 593  
Db 538 TGDTRGANLTISYGGVVDVHKNTLSLGAQGNINITAKQDIAFEKGSNOV-----ITGQ 590  
Qy 594 GTVTITGEGKDFRANNYSLANGTKGLNIISVNN--LTHNLSTINISGNITINOTTRK 650  
Db 591 GTIT-SGNQKGFENNYSLANGTSGLOFTTKRTNKYAITNKFECTNLISGKVNISMVLPK 649  
Qy 651 NTS-YWQTSDDSHWNVSALNLETGANFTF-IKYISSNSKGLITTOYRSSAGVNFNGVNGM 708  
Db 650 NESGYDKPKRTYWNLTSLNVSSEGEFNLITDSRGSDSAGTLTQPYNLNGISF---NKDT 706  
Qy 709 SFNLKEGAKVNFKLKPNENMTSPLP-IRFLANITATGGGSVFPDIYANHSQ---RGAE 764  
Db 707 TFNVARNARVFDIKAPIGINKYSSLYASPNISVSGGSVDFTLLASSNNVOTPGV 766  
Qy 765 LKMSSEINSGANFTNSHVRGDDAFKLNKDLTINATNSFNLSRQTKDFVDGVARNAIN 824  
Db 767 INSKYFNVSTGSSLRFTSGSTKTFSTKDLTLNATGGNITLLQVEGT--DGMIGKGI 824  
Qy 825 STYNISILGGNVTLGGONSSITGNITIEKAAVNTLEANNAPNOQNRORVILKGLSV 884  
Db 825 AKKNITFEGGNITGSKRAVTEIGENVYINNNANVTLIGSDFDNHQ--KPLTIKKVLIIN 882  
Qy 885 NCSLTCENADIKGNLTISISATFKGTRTLNITGNFTNNGTAENITQGVVKLGAVT 944  
Db 883 SGNLTAGGNIVNIAGNLTVESNAFKATINFTFNVGGLFDNKGNSNISIAKGGARFKDID 942  
Qy 945 NQDGLNITHAKRNORSIIGDINDKGLNITDSNNDAEIQIGNISQKEGNLTISDDK 1004  
Db 943 NKSLSITNSSTYRTIISGNITNKGDLNITNBSGSDTEMQIGDVSQKEGNLTISDDK 1002  
Qy 1005 INITKOITIKKIDGEDSSDATSANLITIKTELKLPEDLSISGFNKAETAKDGRDLT 1064  
Db 1003 INITKOITIKAGVDGENSDATNANLITIKTELKLPEDLSISGFNKAETAKDGRDLT 1062  
Qy 1065 IGNSNDGSGABAKVTFTFNNVKDSKISADGHNVTILNSKVKTSSSNGGRESNDDTGLTI 1124  
Db 1063 IGTNTSAD-GTNAKVTFNQVVKDSKISADGHNVTILNSKVKTSSSNGGRESNDDTGLTI 1121  
Qy 1125 TAKNVEVNDITSLKTVNITA-SEKVTITTAGSTINATNGKASITTKT----- 1170  
Db 1122 DAKNVTNNNITSHKAVSISATSGEITTKTGTINATTGNVEITAQTSILGGIBSSSGS 1181  
Qy 1171 -----GDIISGTFISGNT 1181  
Db 1182 VTLTATEGALAVNISGNVTVTANSALTTLAGSTIKGTESVTTSSQSGDITGGTISGCT 1241  
Qy 1182 VSVATVDLTTSKSGKIEAKSGEAVNTSATGTIGTISGNTVNTANAGDLTVGNCAEIN 1241  
Db 1242 VEVKATESLITGNSKIKATTEAVNTSATGTIGTISGNTVNTANAGDLTVGNCAEIN 1301  
Qy 1242 ATEGAATLTFATGNTLTTEAGSSITSTKQGVDDLLAQNQSGTAGSINAVNTLTNTGLTTVA 1301  
Db 1302 ATEGAATLTTSSGKLTTEASSHITSKAGQVNLISAQDGSVAGSINAANVTLTNTGLTTIVK 1361

Qy 1302 GSDIKATSGTTLVINAKDAKLANGSDSTEVNAVNASGSGSVTAATSSSVNITGDLNTVN 1361  
Db 1362 GSNINATSGTTLVINAKDAELNHTVNAVNASGSGSVTAATSSSVNITGDLNIN 1421  
Qy 1362 GSNISKGRNTVRLRGKEIEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEERETLAKLG 1421  
Db 1422 GLNIISKNGINTVLLKGVKIDVKYIQPGIASVDEVEIAKRVLEKVKDLSDEERETLAKLG 1481  
Qy 1422 VSAVRFEVNNITVTNQTNEFTTRPSOVIISSEGRACFSSGNGARVCTNVADGQ 1476  
Db 1482 VSAVRFEIENNTITVDTQNEFATRLSRIVISEGRACFSSGNGARVCTNVADGQ 1536  
RESULT 3  
Q9RNI2 PRELIMINARY; PRT: 1557 AA.  
ID Q9RNI2  
AC Q9RNI2  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)  
DE HmWA.  
GN Haemophilus influenzae.  
OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A950006;  
RX MEDLINE=20359363; PubMed=10899870;  
RA van Schilfgearde M., van Ulsen P., Eijk P., Brand M., Stam M.,  
Kouame J., van Alphen L., Dankert J.;  
RT "Characterization of adherence of nontypeable Haemophilus influenzae  
to human epithelial cells.";  
RL Infect. Immun. 68:4658-4665(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A950006;  
RA van Schilfgearde M., Eijk P., Van Ulsen P.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF180944; AAD56660.1; -;  
DR InterPro: IPR000169; SHprot.acsite.  
DR PROSITE: PS00639; THIOLEPROTEASE HIS; UNKNOWN\_1.  
SQ SEQUENCE 1557 AA; 160461 MW; 62524D384398CB37 CRC64;  
Query Match 63.6%; Score 4711; DB 2; Length 1557;  
Best Local Similarity 63.3%; Pred. No. 2 3e-146;  
Matches 1003; Conservative 155; Mismatches 290; Indels 136; Gaps 23;  
Qy 1 MNKIYRLKFKRLNALYAVSELARGCDHSTKSGEKPARKMKVRLALKPLSAMLISLGV 60  
Db 1 MNKIYRLKFKRLNALYAVSELARGCDHSTKSGEKPARKMKVRLALKPLSAMLISLGV 60  
Qy 61 SIPOSVLASGLQGMVDVHGTATMVDGNKTIIRNSVDALINWKFQNDQEMVQFOLENN 120  
Db 61 SIPOSVLASGLQGMVDVHGTATMVDGNKTIIRNSVDALINWKFQNDQEMVQFOLENN 120  
Qy 121 NSAVFNRTVNTQISQLKGLDLSNGQVFLINPNTGTTIKDAILNTNGTASTLDISNENIK 180  
Db 121 NSAVFNRTVNTQISQLKGLDLSNGQVFLINPNTGTTIKDAILNTNGTASTLDISNENIK 180  
Qy 181 ARNFTFQTKDALKAEIVNHGLITVKGDSVNLIGGKVKNEGVIISVNGGSSISLAGOKIT 240  
Db 181 ARNFTFQTKDALKAEIVNHGLITVKGDSVNLIGGKVKNEGVIISVNGGSSISLAGOKIT 240  
Qy 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSXKDSGNIV 300  
Db 241 ISDIINPTITYSIAAPENEAVALNLDIFAKGGINVRAATIRNOCKLSADSVSXKDSGNIV 300  
Qy 301 LSAKEGAEIIGVYISAOQOAKGKLMITGDKVTLTKTGAVIDLSGEGGETYLGGERGE 360  
Db 301 LSAKEGAEIIGVYISAOQOAKGKLMITGDKVTLTKTGAVIDLSGEGGETYLGGERGE 360

Qy	361	KGKGIQAKKTSLEKGSTINVSKEKGGGFAIVMGDIALIDGNINAOGSGDIAKTGGFVET	420
Db	361	KGKGIQAKKTSLEKGSTINVSKEKGGGFAIVMGDIALIDGNINAOGK-DIAKTGGFVET	419
Qy	421	SGHDLFIKDNAIVDAKEWLLDFDNWSINAEIDPLFNNTGINDPEPTGTGEASDPKK-NSEL	479
Db	420	SGHYLSIGNDAAEVKEWLLDPONVTI-----SNGDDQSOLKDDRGDSPNKILADN	471
Qy	480	KTTLTNTTINLYLKNAWTMNITASRKLTVNNSINIGSNSHLLHSHKSGRGGVGDIDGIT	539
Db	472	KHTVNNKTLTALAKAGIGVNI SAKKVNVTADINV-HNGTLTLHSE---OGGVEINGDIT	527
Qy	540	S-KGGNLTIVSGGVDVHKNITLDQGFLENITA-ASVAFE-GGNNKARDAANAKIVAQGTV	596
Db	528	SEONGNLTIRAGSWVDVHKNITICTGFLNTTAGGSVAFEKAGDKGRAASDAKIVAQGTV	587
Qy	597	TIITGEGRFRANNVSLNGTGGKGLNLS---SVNNLTHNLSGTINISGNITINQITRKNKTS	653
Db	588	T-AGSGODFFNNVSLNGTGRGLKFTAKNGKGNFSAKEDGVNLISGNISINHTANNOLS	646
Qy	654	YWOTSHDSHWNVSALNLETGANF--TFIFK-----YISSNSKGLTTOYRSSAGVNFNGV	704
Db	647	YFHROGYTYWNLTLQNLVNDSDSSFSLTSIKDAIKVGGYDNAKDK-----KNTGGIGF---	697
Qy	705	NGMNSFNLKEGAKVNEFKL-PHENMNTSKPLIRFLANITATGGSV---PFDIVANHSG	760
Db	698	TRDTIENFKOGARKDLSYTLPISPVKNSRIAAVFDGNITVKGGGVNLKFNALSNNYKT	757
Qy	761	RGAEKMKSEINTSGANFTLNSHVRGDDAFKINKDLTINATNSFSLRQTKDDFDVGVAR	820
Db	758	PGVNISRRFINTEGSQLNITCSMPSTTLFNVANDLIINATNSFVSIKEIGT--DHLDD	815
Qy	821	NAINSTYNLISLGGVTLGGQNSSSITGNITIEKAANVTLEANNAPNOQNI RDRVIKLG	880
Db	816	TGLUVNGVNTIKGGNVTILGNSKAKTKFDKNVTVERKANLTILASANFGNH---KGALT	872
Qy	881	SLLVNGSLTLTGENADIKGNLTISESATFKGKTRDTLNI GTNFTNNGTFAEINITGGVVKL	940
Db	873	NINTQKLVATGDTIDVSGDFTVGNDATFNGTNNNLNITGNFTNGTSLIDVKKGAKL	932
Qy	941	GNVYNDGDLNITHAKRNORSIIIGDILNKKGLNITDSNDAETIQGNTISQKEGULTI	1000
Db	933	GNITNPGSUNITHANTNQKTLITGNITNKKGDLNIRDNKNAETIQGNTISQKEGULTI	992
Qy	1001	SSDKINITKQITIKKGIDGEDSSDATSNANLTIKTKELKLTEDLSIGFNKAEITAKDG	1060
Db	993	SSDKVNITKOITIKAGVNGENS DSGTENNANLTIKTKLTLELTNNLISGPHKAEITAKDN	1052
Qy	1061	RLDTIGNSDNGSGAEAKVTFFNVVWVSKSISADGHNVITLNSKVITSSNGGRESND-ND	1119
Db	1053	SDLITIGKASSDSGNAGA QKVPDKVDSKISAGHNHVTLNSEVETSSNGNSNAAGDSNGN	1112
Qy	1120	TGLTITAKNVEYKNDTISLTKWNITASE-KYTTTAGSTPINATNCKASITTKT-----	1170
Db	1113	AGLTISAKDVANNITSHKNTINISATGNTVTTKEGTTINATTTGGEVETARTGDIKGGIE	1172
Qy	1171	-----GDISGT	1176
Db	1173	SKSGGVTLTATGDTILAVGNISGNTVSVTANSGLTLTKADSTIKGTGSVTTLSQSGDIGT	1232
Qy	1177	ISGNTVSVSATVD-LTTKSGSKIEAKSCEANVT-----SATG	1212
Db	1233	ISGKTVSVTATDLSLTVKGGAKINATEGTATLTAASSGKLTTEANSATSGANGVTASSQSG	1292
Qy	1213	TIGGTFISGNTVNTANAGDLTVNGCAEINATEGAATLTATGNTLTIEAGSSITSTKGQVD	1272
Db	1293	DISGTFISGKTVSVTATDSLTVKGGAKINATEGTATLTASSGKLTTEASSITSAKQVD	1352
Qy	1273	LQAQNGSLAGSINANVNTLTGTLTVVAGSDIKATSGTLVINAKDAKNGDASGDSTEV	1332
Db	1353	LSARDGNITGGSINANVNTLTGTLTVKGGSSINANSGLTIVINAEAKDLDTASGRTTV	1412

QY	1333	NAVNAGSGSVTAATSSSVNITGDLNTVRLRKGAEIVKVIQPGVAS	1392
Db	1413	NAVNAGSGSVTAATSSSVNITGDLSTINGLNIISKNGKNTVVLKGAETDVKVIQPGVAS	1472
QY	1393	VEEVTEAKRVLEKVKDLSDEERETLAKLGVSARFVEPNNTTNTVNTQNEFTTRPSSQVTI	1452
Db	1473	AEEVTEAKRALEKVKDLSDEERETLAKLGVSARFVEPNNAITVNTQNEFTTRPSSQVTI	1532
QY	1453	SEGKACFSGGNGARVCTNNVADDQ	1476
Db	1533	SEGKACFSGGNGAAVCTNNVADDGQ	1556
RESULT 4			
Q93DC7			
ID	Q93DC7	PRELIMINARY;	PRT; 2065 AA.
AC	Q93DC7		
DT	01-DEC-2001	(TrEMBLrel. 19, Created)	
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)	
DT	01-MAR-2002	(TrEMBLrel. 20, Last annotation update)	
DE	GN	RSCA.	
GN	RSCA.		
OS	Yersinia enterocolitica.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Yersinia.		
OX	NCBI_TaxID=630;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21437624; PubMed=11533561;		
RA	Nelson K.M., Young G.M., Miller V.L.;		
RT	"Identification of a Locus Involved in Systemic Dissemination of		
RT	Yersinia enterocolitica.;"		
RL	Infect. Immun. 69:6201-6208(2001).		
DR	EMBL: AF394927; AAK77860.1; -		
DR	InterPro: IPR001950; TIF-SUI1.		
DR	PROSITE: PS01118; SUI1_1; UNKNOWN_1.		
SQ	SEQUENCE 2065 AA; 210024 MW; A306D0763682D176 CRC64;		

```

Query Match      12.2%; Score 900; DB 2; Length 2065;
Best Local Similarity 25.0%; Pred. No. 7.8e-22;
Matches 420; Conservative 258; Mismatches 623; Indels 378; Gaps 81;

QY      2 NKIYRLKFSKRLNALVAVSELARG-----DHSTKGSEKPARMKVRHLALK 48
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      3 SKLYKLIFCRRLGLIAGEFTTRYGRSFSFGKKIINDNTRAG-----KLSHLAI- 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY     49 PLSAMLLSLGVTSPQSVLAS---GLOGMVDVHGTATMQVDGNK-TIIRNSVDIAIINWKQ 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     55 -----LTGLALGTLPLVFAHPSPVANG-NITVVGQGMVDNNTTTLTITQSDKLAINWGS 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    105 FNIDQNEWVQLOBNNSAVFNRTVTSNQISOLKILDSNGOVFLINPGLITIGKDAIINT 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    109 FDIAGGNVIYVPGQOSIALNQVLGRDASQIYGNLKRANGQVFLINPGLITFGKAQVDV 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    165 NGFTASTLIDISNENIKARNFTFEOTKOKALAEIYNHGLITVKGDSVNLGGKYKNE--G 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    169 GGLIASTKSMNQDFISGCTYL--TSQKQEGKLVNQLNRLTTAGTGYIALIGQOVDNQPSG 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    223 VISVNGSGISLLAGOKTIS----DIINPTITYSIAPEANEAVNL-----CDIFAKGNI 273
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    227 VINTPQKVALAGSRVILNDRGNLLGVQV-----QGEQVNTLQNGGLIRADSGVI 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    274 NVRA-----ATIRNOGKLSADSVSKDKSGNIVLS-AKEGEAEIGGVISAOQNQAQKG 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    280 QLTAAQKQEMLMNTVIDNTGTLQARGLS-EKNGVITYLNGNGEVVVSQQGMINVNSQQGRGG 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    325 KLMTIGDKVTLKTAVIDLSGKE-GGGETYLGDBERGKNGCIOLAKKTSLEKGTINVSG 383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    339 NVILIGENIHLVAASKIDARDEGGKVLGGDQWQGNKLU-INKRNSVMDKGANIDVSS 397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    384 KEK--CGFATVWGD-IALIDGNINAOQSGDIAKTGTFVETSGHDLF-----IKDNAIV-D 434
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    398 THGACGATVLESHYTGFGYDTHARG-GSLSDGGGVETSSORNLOSFQFVDVSAIMGN 456
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 435 AKEWLLDFONVSI---NAEDPLFNNTGINDERPTG---TGEASDPKKNSELKTLTNTT 488  
Db 457 TGRWLLDPAEVNIVGSAESGVSVQIG---DIPAGVVKNAQVFTPMAN---VTQILNTSI 510  
QY 489 SNYLNKAWNTNITASR---KLTWNSSIN---IGSNHSHLHSGKRGQGGVQIDGD 537  
Db 511 NAQLDKGTNTVITITSSNGSLTNCRCWNTITQADINKTAGGDAYTLT---QADGN 560  
QY 538 ITSCKGNLTIYSGWVDVHNKNTIDOG-----FLNITASVAFEGC-----NNKA 582  
Db 561 IVSNGHNITATG---KUNLLNLLSGDSIVDSIITLNNSDVLLNGDGLLKHANENNA 615  
QY 583 RDAANAKIVAQGVTTITGEGKDFRANKVSLNG-TGK-----GLNISSVNNLHNLSGTIN 637  
Db 616 R-----ISIMGROYGVNLTLEGNTGMAVGVVNSNSAN---ISVAGETR 658  
QY 638 ISG-NITING-----TTRKNTSYWOTSHD-SHWNVSALE---LETGANFTFI 679  
Db 659 ISGESSNANQGWRGIDISGDSVFAGKGNMSFTMTNSRSSWMTFTNATAGDKNITF- 717  
QY 680 KYISSKSGLTTOYRSSAGYNF-NG-----VNGNMSFNK-----EGAKVNF 721  
Db 718 ---QANANGST-----SGGVDFNTGSLVSKSGNISFDINGEIIITOTSFGRLRQGSVSGN 769  
QY 722 LKPNENMTSKPLPIREL---ANITATGGGVSVFDDIVANHSRGAEKMS---EINISNG 775  
Db 770 ---NVNVEINTKGVDFELLRDSHITAG-----NISANATTHKGLWISGDDTDLNASKN 821  
QY 776 ANF---TLNSHVRGDDAFKI---NKDLTIN-ATNSNFSLRQTKDDFDYDGYARNAINSTY- 827  
Db 822 IKLOGVTNSTVTCADAIKISGSSSVQVMAAGNISMAVNVKGTVEG---STISADYT 878  
QY 828 NISILGNGVTIGGONSSSITGNITIEKANVTLEANNAPQOINRIVIKLGSLLVNGS 887  
Db 879 MIKAQGGDFNLISGVKSGPPSNVNI-SADNIIILNGNISON-----DAVV-----MINTF 927  
QY 888 LSLGENADIKNITIS---ESATFKG-----KTRDTLNTGNTFNNGTAENINIGGVVKLG 941  
Db 928 LAAG---DIKANLSSPNYKALYFRNGGMSAGQNLVAVNTSGASVEANTITGIANRM 984  
QY 942 NVTDNDGLNITHAKRNORSI-IGDIIIN---KKGSLNITDSNDAEIQGGNISQEGN 997  
Db 985 NVTVGRDISIIANNHGNMTGAGIGVYVNFPEAKNG--NFTANN-----GSKSIGIAN 1035  
QY 998 LTISDKINIKQITIKKI---DGEDSSDATSNANITIKTELKLTEDLSISGFNAE 1054  
Db 1036 ADIRANEVNLISNTSRADGVVIREANIITLTGNINANATSNRGIVIRONTLSAQKELM 1095  
QY 1055 ITAKDGRDLITGNSNDGSGAEAKTVTFNNYKD---SKISADGHNVTLSKVKTSSSNGG 1111  
Db 1096 LAA-----TSSAASEAIIVOGLSDERSHLVAQG-----NISLKGNSKGS 1136  
QY 1112 RESNSDN---DTGLTITAKNVEVNDKITSILKTVNITASEKVTITAGTINATNGKASITTK 1169  
Db 1137 NPRSSVNLNLSVTSVGNKIDIN-----SSS 1162  
QY 1170 TGDISTGSGNTVSVSAVDTLTGSKSIEAKSGEANVTATG-----TIGGTISGNTV 1223  
Db 1163 VGD---GDYFNNVDLNAALGNVTVYGEALSALSTATNSVLGGNNSIKALNGLICKAI 1220  
QY 1224 NYTANAGDLTVNGAEINATEGAATLTATNLTITTEACSSI-----TSKGVQVDLIAQ 1276  
Db 1221 NTSQAGAGTLFRANGS-LSVAGNIAIQGEGTGATNGIATFYGANTLNIAKDSQLLGE 1279  
QY 1277 N-----GSTAGSINAANVTFLNTTGT-----TTVAGSDIKATSTIVINAK-DAKL 1321  
Db 1280 NTGSOITAGNGISYLSPLKTLTINNNGSLTMBGRSTSGAGINFTPGNNTVVLNGBGDSLI 1339  
QY 1322 NGDA-SGDSSTEVNAVNASGGSVTAATSSS-----VNITDGLNTVNLN 1364  
Db 1340 KGSSVAGSGVAISGMVNNSSGPTIEGSGTSDGSGVHLFSAEHQINRINVTGSSQAEGLR 1399

QY 1365 IISKDRNTVRLKRIEIVKYIQFVSVASVEEVIEAKRVLEKVKDLSDEERTLAKLGVSA 1424  
Db 1400 ISGNATITDALSGRKSINGSGIK-----VDSLLSGSVITHVVDLNAALNGSTHGIGVEI 1454  
QY 1425 VRFVSP-----NNT-----ITVNTQNEFTTRPSSOVIISSEKACFSSSGARVCTN 1470  
Db 1455 TSDINGIHQSIINGTTDGIYGDIDKKNLHVGTSETDILLTQGVATTGSGTGKLDGN 1513  
RESULT 5  
QY 082BY3 PRELIMINARY: PRT: 1910 AA.  
AC 082BY3:  
DT 01-MAR-2002 (TremBrel. 20, Created)  
DT 01-MAR-2002 (TremBrel. 20, Last sequence update)  
DT 01-MAR-2002 (TremBrel. 20, Last annotation update)  
DE putative adhesin.  
GN HMMR OR YFO3247.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Felwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford B.G.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,  
RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
RL Nature 413:523-527(2001).  
DR EMBL; AJ414156; CAC92482.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1910 AA: 195076 MW: 65E5127A098E4045 CRC64;  
Query Match 11.8%; Score 876; DB 16; Length 1910;  
Best Local Similarity 25.5%; Pred. No. 4.3e-21;  
Matches 421; Conservative 235; Mismatches 665; Indels 332; Gaps 79;  
QY 2 NKTYRLKFSKRNALVAVSELRGCDHSTKSGSEKPAKVRHLKLPKLSAMLLSGVTS 61  
Db 3 SKLYKLIFRCRLCLAVGEFTSYGRFSSKGGQAGNORRAVGILSLAMWTGLALGI 62  
QY 62 IPOSVIASGLQGM---VHGTATMVDGDKNTIIRNSVDAT-INNKOFNIDONEMVQFLOE 118  
Db 63 FPLLVLAHPVLPVNGHVVIQGMDDQSSLTVTQOTDKLAINWDSFDIAHGSVIYAQP 122  
QY 119 NNSAVFNRTVSNQISOLKIDNSGVFLNPNNGITIGKDAINTNGFTASTLDISNEN 178  
Db 123 GSOSIALNQOGSASQIYGRQLANGQVFLNPRGILFGKEAQNQVGLVASTKYMSNPE 182  
QY 179 IKARNEFTFEOTKALAEIYNHGLITVGKDSVNLIGGKVKNE--GVISVNGSGISLLAG 236  
Db 183 FLSGDYRL--IGGESEGNIIINQANLRSAPGGYIALVGNRIDNQRSGSITTPQGTNVLAVG 240  
QY 237 QKITIS----DIINPTITYSIAAPENAV-----NLGDIFAKGGINVRA-----AT 279  
Db 241 HSVTLNLDHGNLLGVQI-----QGETVAALIQNGGLIQADGVQIQLTAKGMDLMDTV 293  
QY 280 IRNOGKLSADSVSKDSKSNIVL-SAKEGEAEIGVISAQNOQAQGGKMITGDKVTLKTG 338  
Db 294 IDNTGILQAKGLSA-KNGAIIYDGGEGVVVSQMGITDVNNQOGRGAVVEGKRIVLNKN 352  
QY 339 AVIDLSKEGETYL-CGDERGEGKNGIOAKKTSLEKSTINVSCKE--GGFAIVWG-394  
Db 353 SNIAKTAGGTGVLVGGGQWQK--DNQIRNATAVMDKGSIDIVSASRNGPGGSVAVLWSE 411  
QY 395 DIALIDGINNAQG---SGDIATKGTGFVETSGH----DLFIKDNAIV---DAKEHLLDFDNV 445

Db 412 DYTGFHGNTRARGGQSGD-----GGQVETSSQRNLQAQGVQDASAVRGSAGYWLDDPAEV 467  
QY 446 SI---NAEDPLFNNTG- INDEFPTGTGEADPKKNSLKTTLTNTTISLYLKNAMTMIT 501  
Db 468 TVSSGAESGVMTKVGNIAPAEF- FSSAHIFIPANI---TQILNSINTQLNSGNTVIT 523  
QY 502 ASRKLIVNSINIGNSHLILSHKQGGGQVGDGDIYSKGG---NLTIYSGWVDVHKH 558  
Db 524 TS-----NSSLT-----GCQMCNITVQADITKTAGADATLTLQADGNIVVNN 566  
QY 559 ITLDQGLNITAASVAFEGGNKARDAA---NAKVAQGTVTITGEGKDFRANVNSLNGT 615  
Db 567 ITADAGKLNLLA-----GNTTADSAITLNNKVVLL-----NGDFLAKHANDNT 613  
QY 616 GGLNLITIS---SVNLTNLSSTINISGNITINOTTRKNTSYWQTSKSHWNVSALNLET 672  
Db 614 AR-IGLLGRYDVGNFT---LDGNTALASQGVGNISNAANISV----- 652  
QY 673 GANFTFIKYSISNSKGLTQYRSSAGVNFN-----GVNGNSFNLEKAGKVNFKLPNE 726  
Db 653 -AGEIVISGVNSRG-----QGWKGIDISNNSIULTGV-GNMTFSIGNSNVSW-MGAF 704  
QY 727 NNTSKPLPIRFLANITATGG-----GSVFDDI-----YANHSRG 762  
Db 705 NATITSDKNIIPOGTGSSGGVDFVNSRILSKGRVLEDINGNIVKVVGLRVNSQLS 764  
QY 763 AELKKEINISGANTFL-NSHV---RGD---DAFKINKDLTINA-TNSNFSLR-----Q 809  
Db 765 ADKVFANVTVDGFLRDHSHTVATSGDINANANTINKGIWISGKTNLASGNVNLHG 824  
QY 810 TKDDFYDGYARNAINSTYISITLGNVTGLGONSSSITG-----LMTNTELTAKG---DIK 849  
Db 825 TTNSAYAG-ADAIKISSSSNNVNITAGHISLIANGGKEIGSTVSDVDYANIKN 882  
QY 850 -----NITTEKAA---NVYLEANNAPNOONI---RDRVTKLGSLLVNGSLTGENADIKGN 900  
Db 883 DFNLNITMKSGSPNNATITANNISMGNTITANDAV-----LMTNTELTAKG---DIK 934  
QY 901 LTI-SESATFKGK-----TRDPLNTIGTNGTAEINITQGVKVLGNVTNDGDLNITHA 955  
Db 935 LTPSKGLWFRNGGWTANLILVANSTSGE-----TVKI-NASSSNKNMITA-- 983  
QY 956 KRNORSITGGDIKNKG-SLNTDSN---NDAEIQIGNISQKEGNLTISSDKINIKOI 1011  
Db 984 -GKDISIAGNSKATGPNIENIENVTNNGFTNGITSTWLSGVNVSANGVDITSNS 1042  
QY 1012 TIKKGIDGEDSSDATSNANLTIKTELK-LTEDLSIGFNKAEITAKGRDLT-IGNS 1068  
Db 1043 TGTGGI-----VLDTNLTITVGDINTIVTNSGKGIWIKSNSTLNSNKDITLVGS 1094  
QY 1069 NDGSGAEPAKTVTFNNYKD---SKISADGHNVTILNSKVKTSSNGGREGNSDN--DTGLT 1123  
Db 1095 AGQNEG-----VITQSSDASRNNISAQG-NITLIGKM-----GNGSCQHSLLNLGNVSLT 1144  
QY 1124 ITAKNVEVKNKDI5LKVNTITASEKVTTTAGSTINATNGKASITTKTGDISGNTVS 1183  
Db 1145 SSGRNDIDINGSAGTGVDVFTNVE-----LNATAGNVSIYAE-----NTALS 1187  
QY 1184 VSATVDTLTKSGKIEAKSEANVTATGTIGTISGNTVNTANAGDLTVNGAEINAT 1243  
Db 1188 TSLNAVLSLGNNSIKAQNG-----WLIGKAFNTQAGIGFRANSSLSVDGNIIKGET 1242  
QY 1244 EGAATLFTAG-----NLTTFEAGSSIT---STKGVQDVLIAQNGSTAGSINAANVTNLT 1294  
Db 1243 EGVGA-TRKIDFYGANTLNIKGSQSLGKNGKAQDTAGNGISYTSI--AKLIVNN 1299  
QY 1295 GTL-----TTVAGSDIKATSGTLVYNAK--DAKLGDA-SGDSTEYNAVNASGGSVTA 1346  
Db 1300 GSKMEGRSTSGTGINFPSSNNLTFLVNGDGTLLIKGSSVAGTGAATISGVNNSTGPM 1359  
QY 1347 TSSS-----VNITGDLNTVNGLNIISKDRNTVRLRGKEIEVKYIQ--- 1387  
Db 1360 GISTDAGVHLFSAEHRIDRINTVSGSTHAEGLRISGNAIIVDTTLTGKSGNSGVKIDS 1419

QY 1388 -PGVASVEEVTEAKRVLEK--VKDLSDEERTLAKLGVSAVRFVEPNNTIT-----VN 1437  
Db 1420 LPSGSV-----TRSYLDNATLNGSSSGKGVETISDINGIHSSINGTTGTGCGVIG 1474  
QY 1438 TONEFTTRPSQVITISEGKACFSSGNGARVCTN 1470  
Db 1475 ENSNVTGTSEADLLILOQVATTGTGTGIKNGN 1507

## RESULT 6

ID Q8XQ42 PRELIMINARY; PRT; 1371 AA.  
AC Q8XQ42;  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Putative hemagglutinin-related protein.  
GN RSP1444 OR RS03099.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OG Plasmid megaplasmid.  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GMI1000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,  
Weissenbach J., Boucher C.A.  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."  
RL Nature 415:497-502(2002).  
DR EMBL; AL646084; CAD18595.1; -  
DR InterPro; IPR002064; DNA\_pol\_B.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; UNKNOWN\_1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 1371 AA; 132558 MW; 1D9AA3037BF8CF8 CRC64;

Query Match 10.6%; Score 785; DB 16; Length 1371;  
Best Local Similarity 24.2%; Pred. No. 2.8e-18;  
Matches 365; Conservative 196; Mismatches 533; Indels 414; Gaps 61;

QY 42 VRH-----LALKPLSAMLLSLGVTs---IPQSVLASLQGM----- 74.  
Db 1 MRHFTLHVALIPSGAAITRAAVTGHARHPRAPRNALGGTVLLGTFAAMPFAAPAG 60  
QY 75 -DVVHGATMQVDG-NKTIIRNSVDAILNKOFIDQEMVQFLQENNNNSAVFNRTSNQ 132  
Db 61 GTVTSGSGSIGONGTITITOTSSRLAIDWSAFQAGETVNFIQPGAGAIALNRVTGHE 120  
QY 133 ISOLGILDSNGQVELINPNGITICKDALIINTNGTASTLDISNENIKARFTFQTKDK 192  
Db 121 ATSLGSLNANGTVFLLPNPVGFGAGVNVGVGLVASTLGLSNADFEAGRYAL---SGG 177  
QY 193 ALAEIVNHGLITVKGDSVNLIGKVKNEGVISVNGGSLSLAGOKITISDIINPTIYS 252  
Db 178 STAGVTNCGTITVPSGKVALIANSDVNGNISAPRGVSLAGAGNVTLTLDGSPLYGT 237  
QY 253 IAAPNEAV--NLGDIKAFKGN-----NVRAAIT-RNQG--KLSA 288  
Db 238 ISAGAARTLVNNGGMIVADGGRVLTARGLDLSSESVNNTTGVVARTVGNNGQTIELIG 297  
QY 289 DSVS--KDKSGNVLISAKEGAEIGVISAOQKAGKGLMTIGDKVTLTKCAVIDLSGK 346  
Db 298 DPVAGLTQVSGQIDASAPDG-----GG-----NGGSVKVLGTVGVFGAIDVSGM 344  
QY 347 EGGETYL--GGDERGCKNGIQIAKKTSLKSTINVSQ--KEKGGFAIWWG-DIALIDGN 402  
Db 345 AGGCTALLIGGNAQAGPE--PNATATYVAPTAAIDASAIROGNGRLLIVGWTDVANVHGS 402

```
QY 403 INAQSGDIATKGGFVETSGHDLFIKDNAIVDAKE-----WILDFONVSNIAEDPLFN 455
Db 403 L5ANG-GAOGGNGHGVETSGHALDTGDIASVAGGGGTGGWLLDLPYNVLSA-----455
QY 456 NTGINDPFTGGEASDPKNSSELKTTLTNTTISNYLKNWMTNITAS--RKLTVNSSIN 513
Db 456 -----GTQTCGA-----FSGNVWTPSAGSLVNTSSIQSLN 487
QY 514 IGSNHLILHSKGOORGGOIGDITP-SKGN--LTIYSGGV-----DVHKNITLDOG 564
Db 488 SGNVVTITTTAGTQBGNTAGNSISKTAGGNASLTLIADGRITTNASSGTHRTITTSIG 547
QY 565 FLNITAASVAFEGGNKARDAANAKIVAQGTVTITGEGKDFRANVSLNGTKGLNISS 624
Db 548 TLDVSM-----TARATTSAASTSGINLRYVDINANGNISATASGAQSGTA 593
QY 625 VNNLTHNLSTINISNITINOTTRKNTSYQWTSDDSHWNVSALNLETGANFTFIKYISS 684
Db 594 AALSQNSWSTTGAGNISLSGLTPDNGN---SQGVFLNANTLTITASG-----ITVSG 644
QY 685 NSKGLTQYRSSAGVNFVG---NGNMSFNLKEGAKVNFKLKPNMNMNTSKPLPIRFLA 740
Db 645 TSGCIAT---ITGTSNGAPLFTSTNIGVLAGG-----GSGVTLTSLA---DSLSTSGAVTISG 722
QY 741 NITATGGSGVFEDYANHSRGRGAELKMSINISNGANFTLNSHVRGDDAFKINKDLTINA 800
Db 678 -LRSTGGA---LALTGTATGATATWA-----GSGVTLTSLA---DSLSTSGAVTISG 722
QY 801 TNSNFSLRQTKDDFYDGYARNAINSTY-----NI-----SILGGMVTLGGONS--843
Db 723 T-----ASNPVSSYRNQSLSTVNIACTGNSAASLTGCTVTITGTSNV 765
QY 844 ---SSITGNITTEKAAVNTLEANNAPNOQNIHDKIKGLSLLVNGSLSTGSENADIKG-859
Db 766 GNASTSNGNAVKLDGKNLTATAGP-----INIAAGSAGDGV 805
QY 900 -----NLITSEATFKGTRDTPLN-ITGNFTNNGTAEINI-TQGVVKL-----940
Db 806 WSGSGAVTMSAPASSSITISARLSDSVSGYSGFYIGGGSATLTFATAAPVSAESLVG 865
QY 941 -----GNVNDGDLNI-----TTHAKRNORSIIGDIINKG---SLNITDS 979
Db 866 ARKAFWNKGGTLVPGNLRIVTTGGAIADDTTHG--GYFHTVGTSTSDSGAGNVTSLNS 923
QY 980 NNDABEIQIGGNISQEGNLTISDPKINIKQITIKKIGIDGEDSSDATSANLTIKTKEL 1039
Db 924 GN---TFAGALSINAGDPTLAS-----STSLTL-----DSSTVNGALNLTAP---962
QY 1040 KLTEDSLISGFNKAETAKDGRDLTIGNSDNGSGBAEKVTTFNNVKDSKISADGHNVTL 1099
Db 963 GITOSGAVSAAGTATL---NGNAAITLNDGNA-----FNAL---AVTGAAGVNV--1007
QY 1100 NSKVTSSNGGRESNDSNDTGLTITAKNVEVVKDITSLKTVNITASEKVTITTAGSTINA 1159
Db 1008 -----VNAALSVTGTSTGAVRFTTSLGDLATR 1036
QY 1160 TN-----GKASITTKTGDISGTISGNTVSVSATVDLTTKSGSKIEAKSGEANVTSATGTI 1214
Db 1037 GNWSISGDLTLISAGAGNTRGTASGNNVDSQAT--LAVDAGRTITVYSGAVGSTVLGGTL 1094
QY 1215 GGTISGNTVNTANA--GDLT---VGNCAEINATEGAATLTATGNTLTTEAGSSITSTK 1268
Db 1095 AGRAAGSGNFYNRQDGAFCAGDAGIGDGYVYVYRPTVTI---PTDAANTKVDG 1150
QY 1269 GOVDLLAQNAGSINANANTVNTTGTTLTVAGSD-----1304
Db 1151 GASDPALAYTIVGOVNGDSATQIILSGSLARAAGNACRYAISQGLTADQIGYQTVLDVG 1210
QY 1305 ---IKATSGTLVINAOKAKLNGDASGDSTEVNNAVAGSGSVTAATSSSVNITGDLNIVNG 1362
Db 1211 HAFSITPAPLLITA-----NGASRTAGQDNPAFSATFAGVGGGTQATDLRGLRLTTS 1265
```

```
QY 1363 LNIISKDG 1370
Db 1266 ADAGSPAG 1273
```

## RESULT 7

```
Q9HVG6 PRELIMINARY; PRT; 2154 AA.
AC Q9HVG6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein PA4625.
GN PA4625.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01; PubMed=10984043;
RA MEDLINP=20437337; Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Stover C.K., Pham X.-Q.T., Huynh W.O., Kowalik D.J., Lagrou M.,
RA Hickey M.J., Brinkman F.S.D., Hufnagle W.O., Westbrock-Wadman S., Yuan Y.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Lim R.M.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004876; AAG08013.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2154 AA; 219755 MW; 4B8ABCE66968A0CE CRC64;
```

```
Query Match 10.1%; Score 744.5; DB 16; Length 2154;
Best Local Similarity 23.9%; Pred. No. 9.7e-17;
Matches 394; Conservative 231; Mismatches 585; Indels 437; Gaps 80;
```

```
QY 87 GSKTIIRNSVD-AIINWKQFNIDQEMVQFLOENNNNSAVFNRTSNQISQLKGLDSSNQ 145
Db 60 GKTLLIDQSTQRAIINWKGFVDSADAEVRFNPQVTSSTLNRVTAGQESVIAGRISAPQ 119
QY 146 VELINPENGITIGKDATINTNGTASTLDISNENIKARNFTFEQTKDALKAAEIVNHGLIV 205
Db 120 VIYNSGVVFSKAKVDYVGSLLTITITANISDEHFRQGLIFDQPGNPD-ARIVNDGSISV 178
QY 206 GRDGSVNLIGGKVKNEGVSIVNGSGSISLLAGOKITI-----SDIINPTITYSIAAPENEA- 260
Db 179 AEKGLAFAVPSVANNGVINARLCTVMAAGNAATIDYDGLVSTAVTDPVTRKPODAQ 238
QY 261 ---VNLGDFAFKGNINVRA---ATIRNOGKLSADSVSKD---KSNIVLSAKGEAEI 310
Db 239 ALVNSGAIQADGGSVLITAEQASRVVDNAVNLGVLARGIEVREGSVALVSKSDIOI 298
QY 311 GGVSIAQNOQAKGKLMITGDKVTLTKTGAVIDLSG-KEGETYLLGDEGKNGKOLAK 369
Db 299 AGKIDVSGPK-NGGDVLVSGQVVALASTASIDARGTAQGSVRIGDFQGE--LPRAK 355
QY 370 KTSLEKSGTINVS--GREKGGFAIVNGD-IALIDGINAOGSGDIAKTGGFVETSGH-DL 425
Db 356 NATLAKGASIDVATCKNGGGLAVVWSDGNTRMDGRILARG-GAQQGNGGLVETSKVNL 414
QY 426 FIKDNAIV-----DAKEWLLDFDNVSNIAEDPLFNNTGINDPFTTGEASDPKKNSE 478
Db 415 SIADSAVSVAAAPYNGGTWLLDPTTLRIVA-----SGTSGSVGGANGASDATVNAS 468
QY 479 LKT-TLNTNTTISVYLNKAWTMNTITASKRLTVNSSI---NIGSNH-LIILHSGQGGVQ 533
Db 469 VVTGALAGGRVT-----LSASDRLSVEAPLITSNLGASRGLELIATGP-AGAVD 517
QY 534 IDGDIITSK-----GGLTIYSGWVDVHKHNTLIDQGLF---NITAASVAFEGGNNA 582
```



Db 518 ISAPILFRNGSLAIRAGGNINFLSGGTPQTSQIVDLGSGTLMNQTSTAGKISQOAGT--A 575  
QY 583 RDAAN-----AKIVAQGVTTITGEGKDFRANNVSLNG-----T 615  
Db 576 LIAANLAGRAGSIDLASWDNYAGNALQTFNGTLKYRQSNATGVTSIGVFDFINQSWT 635  
QY 616 KGGLNISSVNN--LTHNLSGTINISGNITINQTTKNTSY-----W 655  
Db 636 GTAQNVSSVGTIRLEANSVGT--TGNYTL--TADGNSEDFRLVFTALPYRRVSGSASF 690  
QY 656 QTSHDHSHVNSALNLE--TGANFTFK-----YISSNSKGLTT---QYRSSAGV-NFN 702  
Db 691 PTNDSDDLVTNLRVQVNGSNVATPNGCAPSGFTVAAGNGSVTTWTGNGTWSGKFGG 750  
QY 703 GYNG-----NMSENLKEGAKVNFKLKPNENN--- 728  
Db 751 VIGVTDLOYDVGTLGTEELIFGLGKTSRVDTRDLFMRECAFNSFAERAQVEMKTT 810  
QY 729 -----NTSKPLRFLAN-----ITATGGSVFDDIYANHSRG---AELKMS 768  
Db 811 TTAGDILSRQQTATLTANDATRVYGDVNPFTLATMSGINADAYVNSQFNLDYQATSTT 870  
QY 769 EINISNGAFTNSHVRGDAFK-----INKDLTINATNSFSLRQTKDDFDVGYARNAI 823  
Db 871 ATQASNVGQYAITGNANGSEYFQRYQLVRQDGLTVPALIVSADAKTKVYGDADPTL 930  
QY 824 NSTYNIS-----ILGGNV--TLGQNSSSSTIGNITIEK-----AANVTI--- 861  
Db 931 --TYQVSGLKNSDTAAGVLSGNLGRVAGEN-----VNGYILQGLGLGNTANYTSLVGN 983  
QY 862 EANNAPNOON--IRDRVILK-----GSLVNGSLS--LTGENADIK 898  
Db 984 DLRIITPAQLNVIADAKTKVYGDLPALTYQVSGLRGDTAGAVLNGGSLSRVAGENGVY 1043  
QY 899 GNLTISESATFKGTRDTLNTIGNTNGTAEINTQGVKL-----GNVTNDGDLNIT 953  
Db 1044 G---INQGLGLVSSNYTLNTOGN-----NLTIITKALLNVIADAKTKVYGDADPALTY 1093  
QY 954 HAKNRQSIIOGDIIN-----KKGSLNITDSNDAEIOIGGNISQEG 996  
Db 1094 QVSGLKGNDTAGAVLNGGSLSRVAGENGVYVINGQGLGILLSANYDLSYQ--GN----- 1145  
QY 997 NLTISSDKINIKQITIKKGDGEDSSDTSANLTIKTELK-----LT----- 1042  
Db 1146 NLTIITKALLNVIADAKTKVYGDADPS-----LTYQVSGLKGNDTAGSILTGGLNRA 1196  
QY 1043 --EDLSISCFNAEITAKDGR--DLTIGNSDG-----NSGAFAKTYTFNNVRDSKI--- 1090  
Db 1197 AGENGVYVINGQDLALNSGNYDLSYOGNLTITKALLNVIADAKTKVYGDADPSLTYQV 1256  
QY 1091 ----SADGHNVTLN--SKVTSSNGGRESNDNDGTITITAKNVEYNKDTISLKTVNIT 1144  
Db 1257 SGLKNGDTAGAVLNGGGLVVRVSGENVGYA IQOGLGL-----VSGNYDL--AYQGNLT 1309  
QY 1145 ASEKVTITTAGSTINATNGKA--SIT-----TKGDTISGTI-----SGNTVSV----- 1184  
Db 1310 ITKALLNVIADAKTKVYGDADPSLTYQVSGLKGNDGDSAGSILTGGLNRAAGENGVYGINQ 1369  
QY 1185 -----SATVDLT-----TKSGSKI--EAKS---GEANVTSAITGTTIGTISGNTVNV 1225  
Db 1370 GDALNSGNYDLSYOGNLTITKALLNVIADAKTKVYGDAD--PSLTYQVSGLKGNDGTAGA 1428  
QY 1226 TANAGDLTVNGAEIN--ATEGAATLTATNTLTTEAGSSITKQGVDLA----- 1275  
Db 1429 VLNCGGLVVRVSGENVGYA IQOGLGLVSGNYDLSYOGNLTITKALLNVIADAKTKVYV 1488  
QY 1276 -----QNGSIAGSINAANVTLLTGTTLTVAGSDTKATSGTLVINAKDAKUN 1322  
Db 1489 DADPSLTYQVSGLKGNDTAGAV-----LN--GGSLSRVAGENV-----VGYINGQGLDALN 1537  
QY 1323 GDASGOSTEVNAVNASGGSVTAATSSSVNITGDLMT-----VNGLN----- 1364  
Db 1538 ---SGN-----YDLSYQGNLTITKALLNVIADAKTKVYGDADPSLTYQVSGLKGNDTA 1588

QY 1365 --IISKDGRNTVRLRGKIEVKYIQPG 1389  
Db 1589 GAVL--NGGGLVVRVSGENVGYA IQOQ 1613

RESULT 8  
QYHVNG  
ID QYHVNG PRELIMINARY; PRT; 1417 AA.  
AC QYHVNG; 2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein PA4541.  
GN PA4541.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
Brody R.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.F.,  
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964 (2000).  
DR EMBL; AE004867; AAC07929.1; --  
DR HSSP; P22629; LSWC.  
DR InterPro; IPR003880; Ppantne\_attach.  
DR PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1417 AA; 139958 MW; 212C916D5A55C39D CRC64;

Query Match  
Best Local Similarity 21.9%; Score 636.5; DB 16; Length 1417;  
Matches 352; Conservative 238; Mismatches 523; Indels 495; Gaps 72;

QY 1 MKKIYRLKFSKRLNALVAVSELRGCDHTEKSGEPARKVRRHLAKPLSAMLISLGYT 60  
Db 1 MNKSYTLVWQ-----ATGCWNVASEGTRRRSK--SGRKAALVAVAGASLLGL-PC 47  
QY 61 SIPOSVLASGLOGMDVVGHTATMQ--VDGNKTLII-RNSVDALINMKQFNIDQNMVQFLQ 117  
Db 48 QAPAFALPS--GATVVSQDAGFQSTDRHVMVIDQSHKLITNNFESVRADERVSFHQ 104  
QY 118 ENNSAVFNRTSNQISQLKILDSNGQVFLNPNGITIGDKAIINTNGFTASTLDSNE 177  
Db 105 PGQDAVALNRVIRNGSDIOGRIDANGKVFVLPNGVVGKSAQVNVGGLVASTLDLDR 164  
QY 178 NIKARNTFEQTKDALAEIVNHGLITVKGDSVNLGKVKNEGVISVNGSGISLLAQ 237  
Db 165 DFLAGNYQFSGSG---ATVSNAGSLQASEGSIALLGARVSDGLIQAQLGDVALGAG 221  
QY 238 KITSDIINPTITYSTIAAPENEAVALNGLDIFAKGGNINRA-----ATIRNOGLSADSV 291  
Db 222 GINLN-----FDGDLNLNLDVQKGSVDLAHNGGLLRAD-- 255  
QY 292 SKDKSGNIVLSAKGEAEIGGVISAQN-----QQAQGGKMLITGDKVTLTKTGAVIDLS 344  
Db 256 ---GGQVLMARSADSLTKTVNNQGTLEARTLSRAGEGRIVLDGEGQGTVRVAGKQDAS 311  
QY 345 GKEGETVLGDERGEGKNGIQIAK---KTSLEKGSITNVSGKGGKGFALVWGDIALID 400  
Db 312 AIGGN---GGLVLNQGAN--VEIORTAQVDTHADGAT-----GTWRLISHEVSVA 359  
QY 401 -GNTNAQSGDIATGGFVETSGHDLFTIKDINAIVDAKEWLLDFDNVNSINAEPLFNNTGI 459

360	VGQANAAGDG-----SQQVHVAAQGPAGANAS-----DSNGVTIVQOQP-----	397
460	NDEFFPTGGEADPKPKNSLKTTLTNTTISYLNKNAWMTITASRKULTVNSNINIGSNSH	519
398	-----AVDLAAGANGTSAVOSOGANIGSANGISVVOQO-----NSPNIGSGAN	442
520	LILHKGQGGGV-----QIDGDTTSKGMGLTIYSGGWVDVHKHNTLDGFLN	567
443	GISVVOQGANIGAGASGISVVOQSPNPGVGVTV-----VOSOQGANIGSGASG	497
568	ITAASYAFEGGNKKARDANAAKIYAQGVTTITGECKDFRANNYSLANGTKGLNIITSSVNN	627
498	ITV-----VOSOQGANIGSGASGISVVOQSPGSGISG-----VNGVTIVOSOQGANIGPGVSG	551
628	L-----THNLSGTINISGNITINQOTTRKNTSYQWTSHD-----SHMNVSAINLETGANFT	677
552	IDVVQOTLPNLPSGANGSSIVQV-----QTLPOIAADAGNVHVQVQVQGTGNGKV	600
678	FIKYISSNSKGLTTQYRSSAGVNFVNGVNMGSFNLKEGAKVNFKLKPENNMNTSKPLPIR	737
601	F-----GNSATNVRSR-----TVQARSNENVGSG-----	624
738	FLANITATGGSVFFDIYANHSGRAELKMKSEINLSGANGFTLNSH-----	784
625	-LANPSAGKGST--LHADTLARNLSITSNVVVYATRG-----NAHVGAPLSWDSNGLGT	675
785	-----RGDDAFKINKDLTINATNSFNSLROTKDDFYDYARNAINSTYNIILGCVNLGG	840
676	LTAERG--LRINGALTAQGENASITLNA-----GORPLRIDDSLSLTCQCARVEF--	724
841	QNSSSSTIGNITTEKAANYTLBANNAPQOONIRDRV-----KLGSLVYNGSL	888
725	-NSDK-----GYALAEGRITLSKGNAGFRANGROYSVTDLOQLRGIDRDLGGSYVLGN-	778
889	SLTGENADIKGNITISEATF-----KGKTRDTLNI--TGNTF-----NNGT-----	928
779	RIAGGNSSF---LSIGNASAFGGTFDGLGNTIDNLAUVGTGAYSGLFSVNGRTLRNLNLE	835
929	-----AEINITOG-----VVKLGNV--TNDGDLNITHAKRNORSIIIGDIIKKGSL	974
836	RISADGAQATHYVWQVGSALAVNLGRIDNVNASDIRIAAASKLNS-----LGLVALNLGSI	892
975	NITDSN-----NDAEIQIG-----NISOKEGNLTISDKKINITKOITIKKGIDGEDSSD	1025
893	DNASAGSLTVGNRHTVALGAAENISTARGVASTNSR-----	931
1026	ATSNANLITIKTELKLTEDLSISGFKBAITAKDGRDLTIGNSNDG-----NSGAEKATVTF	1082
932	-----ADFAISGOLK-DHASYHAGGLVGNRGGILTRSSGO-----	967
1083	NNVKDSKISADGHNVTLNSKVKTSSSNGRESNNDT-----GLTTAKNVENVNKDIT	1136
968	-----GTLSLSGHGMNLGGLVGSAGGLADYSASVDVSGNGRGLYGGLIGLVNMSGI-	1021
1137	SLKTVNITASEKVTTR-----AGSTINATNGKASITTKTGDIS-----GTL-----	1177
1022	---AHATASGVKRGTDAAELGGLIGRNLNAAINNAS--AHGDVSLQAGRYLGLLIGHN	1074
1178	-SGNTVSVSATVDLTTKSGSKIEA-----KSGEANVT-----SATGTIGGT	1217
1075	QAGNLNVVTSYGNLS--GGSLQAGGLIGLGNASLVIVNASKNGVATRGAEVAGGLIGEN	1132
1218	ISGNTVNVNTAGDLTVNGAEINATEGA-----ATLTAT-----	1252
1133	LYGVSWINGSA-SGEVTDGSGKTLGGGLIGLNGLNHNSLNKASGVWVAGANSVDVGLLIGHNR	1191
1253	-GNTLTTEAGSSSTTKGQVDDLLAQNGSIAGSINAANVT-LNMTTGLT-----VAG	1302
1192	GGNHSTLAAASGNVTGGRG-----SRVGGVLGVYNDAAASTNVASNGVSGASRAIGGLIG	1246
1303	SDIKATSGTTLVINAKDAKLNGDAS-----GOSTEVNAVNASGS-----	1340
1247	SDLR---GSLMLASHGHIVNDKTSHLNGLLVGRGENTSRSAKASGVGSGGAGIRAGGIV	1303

QY	1341	-----GSVTAATSSSVN--ITGDLNTVNGLNIIISKDG	1370
		: : :   :   :	
Dh	1304	GSLEGWOALILCASAGGDDTAGYDSYTGGLVGFSTATISGASAGKVG	1351

DECEMBER 9

AD	Q8ZHAL	PRELIMINARY;	PRT;	3705 AA.
AC	Q8ZHAL			
DT	01-MAR-2002	(TEMBLrel. 20, Created)		
DT	01-MAR-2002	(TEMBLrel. 20, Last sequence update)		
DE	01-JUN-2002	(TEMBLrel. 21, Last annotation update)		
DE		putative autotransporter protein.		
GN	YAPH OR YPHO1004.			
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Yersinia.			
OX	NCBI_Taxid=632;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CO-92 / BIOVAR ORIENTALIS;			
RX	MEDLINE=21470413; PubMed=11586360;			
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.C.			
RA	prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.I.			
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.			
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,			
RA	Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,			
RA	Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,			
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.			
RT	"Genome sequence of Yersinia pestis, the causative agent of plague"			
RL	Nature 413:523-527(2001).			
DR	EMBL; AJ414146; CAC89847.1; -			
DR	Interpro; IPR004899; Pectact_sup.			
DR	Pfam; PF03212; Pectactin; 1.			
DR	PRINTS; PR01656; VACCVTOXIN.			
KW	HYPOTHEICAL protein; Complete proteome.			
SQ	SEQUENCE 3705 AA; 370478 MW; 8D36DAB27F85F284 CRC64;			

Query Match	8.1%;	Score 602;	DB 16;	Length 3705;
Best Local Similarity	21.7%;	pred. No. 7.8e-12;		
Matches 413;	Conservative 260;	Mismatches 695;	Indels	534;

Qy	14	NALVAVSFLA--RCGDHSTKGS---EKPARMKVRHLAKPLSLAMLLSLGVTSPISQVSV		
Db	972	SGVTVTDALNSGTVNITGSGSWDNTDPLATNVSILEQDRAGSTLELINATVGTGDI		
Qy	69	SLCGMDVVHCTA-TMQVDGNKTIIRNSVDAILNKKQFNIDQNMVQFLQENNNSAVFV		
Db	1032	LDL-----LVNGTAITSGTQGVSAIQGGSVIA-----NAIHNYGLASSNSNGSDGLY		
Qy	128	VTSNOISQIKGLDSNGOVFLNPNGITCKDAIIN-----		
Db	1083	YT---LSALELLAGADALLATSGLTANR--VLNDELFCVGLVVDQAONGALTLAN		
Qy	165	NGFTASTLIDISNENIKARNFTFQTK-----DKALAEIVNHGLITVTKG		
Db	1138	NRYEQTITVTAGELLGANGAFGOTSLDDIAGSASANINGYSQVGVATWNGVTGSLGS		
Qy	210	SYNLIGGKVKNEGVIYVNGSGISLLAGQKIPISDIINPTITYSTAAPENAEVNLGDIF		
Db	1198	V--LTSGLLTNGGILDLTGGALNITAGASTVAGGLTGAGTLNI-----		
Qy	270	GGNIINVRAATIRNQGKLS-AD--SVSKDKSGNIYVLSAKE--GEAIEGGVISAQNOQAK		
Db	1241	GCNLSVSAANSGLSGOOTHIAOVASVTTDTCTGLTSAVEVLGTLNLNGAAMTN		
Qy	325	KLMITGDKVTKTKTGAVIDLSGKE--GGETYLGGBD---GEGK-----		
Db	1296	--VLSDGQ-TINTNAVTLTSGNNSPSGAHQICTDGLTGVQASNLGASSATVNLGLT		
Qy	363	---NGT--OLAKKTSLEKGSTINVSQ-----KEKGFAIIVGWDIALI-----		

```
Db 1353 LILNGVSESIANVLGVAGSTVDIIGADTALTANNSGFL---GOYALAGNSKLTVASTN 1409
QY 402 NINAQ-----GSGDIAGTGGFVETSHD-----LFIKONAIVD-----AKEWLLDF 442
Db 1410 NLGASSVALAGAGDTLSLGSFNGTSGNSVTGSGVLQVTDAAEVTLTSSNGSVNAVYDI 1469
QY 443 DNVSINAD-PLFN-----NTGIND-----EFPPTGCEASDPKKKSELKTLTNTT 487
Db 1470 ADATLNDLDDALFNHVLGTGNGLLVAKNDASTAFDFGTGGAFSGIVN-----LNTT 1523
QY 488 ISNLYKNAMTMINITASRLKLTWNSSINIGSNHLLHSGKQGGGVQIDG-----536
Db 1524 FALSADNAAL-ARATLKSDSDSVTVTGAYDR-TLHGLDUNGDTLIFDGPQPOANGV 1581
QY 537 ---DITSKGNNTIY-SGGWVDVH-----KNITPL-----DQGFNLITAAAFEGGNKARDA 585
Db 1582 TVTDLALNSGTISITGAGNWEHPVTPPNVSLLEQDRGDILLELINAANTGN-----A 1636
QY 586 ANAKIVAQGVTVITG-----EGKDFRANV-----SLNGTG-----616
Db 1637 NLDDLVDGTATSTGTQGVESAIOQGGSTVANAHINYGLTSSNGNGSGGLXYNYTLSALE 1696
QY 617 ---KGLN---IISVNNLTHNL-----SGTINISGNITINQTRKNKTSYWTSHDSHWYSAL 668
Db 1697 LLANGANALLATESGLTANRVLNABELFCVGLVYDQAQNGALTLANGNRRYEGTITVAG 1756
QY 669 NLEGTANFTF-----IKYISSNKGTLTYQVRSAGVNFNG-----VNGNM 708
Db 1757 ELILGANGAFGQTSLLNIASGASANINGRYQTGVAVTNSGAVTLGNGVLTSGLLTNGGI 1816
QY 709 -----SFNLKEGKRVNF-----KLKPN-----ENMNTSKPLPIREFLANITATG 746
Db 1817 LDLTGGLNALAAGSSVAGLGTAGTGLNNGDLAVSATNSGLSGQTHIADVASVTLTG 1876
QY 747 GGSVFEDYANISRGAEELKMSSEINSGANFTLNSHVRGDDAFKINKDLTINATNSFS 806
Db 1877 TGTL-----GTSAVEVLGTNL-NGANAAMTNVLSGGVINTNAAVTLCGNSFS 1926
QY 807 LRO--TKDFVDGVARN--AINSTYNIISLGGVNTLGGQNS-----SSYGN-ITIEKA 856
Db 1927 AHQIGTDELGVQASNLGASSATVNLGTLTSHLILNGVSESIANVLGVAGSTVDIIG 1986
QY 857 ANYTEANNAPNOONIRDRVIKIGLSLYNGSLTGENADIKNLTISESATIFKGTROT 916
Db 1987 ADTALTANSG-----FLGQVALAGNSKLTVASIN---NLGASSVALAG-AGT 2032
QY 917 LNTGNTNCTABINIT-QGVVKLCNTNDCDLNITP-----HAKNRQSI 962
Db 2033 LSLUG---FNGTFSNVTSGLVQ---VTDDAEVTLTSSNGVNTVKVADIADATLNDI 2086
QY 963 IGGD-IINKKGLNITDSNNDAEIQIG-----GNISQKEGNTLSSDKINTKQITI 1013
Db 2087 ALFDHVLGTGNLVAKNLATTAFDGTGAFSGIVNLNITTFALSADNAALARATL 2146
QY 1014 KKGID-----GED-----SSSDATSNANLTIKTKELKLT-E-LSISGFN 1051
Db 2147 KLSDDSVTVTGTDRILHGLDLNGTLIFDGPQPOANGVTVTDLALNSGTISITGAG 2206
QY 1052 KAB-----ITAKGDRDLTI-----GNSDNGSGAEAKTVTFNVVKSKIS 1091
Db 2207 WNENEHPVTPPNVSLLEQDRGDILLELIDADNVTGNANLEMLINGTISAGOGVOSTVQ 2266
QY 1092 AGHNVTLNSKVTSSNGR-----ESND-----NDTGLTIPAKKVEVN 1132
Db 2267 QGGYTVANATHYGMT-SNGSGSLYVYVNTLSALELLADGANALLATESGLT---ANRELN 2323
QY 1133 KDISLKTVNTITASE-----KVTTAGSTTNATNKASITTKTODISGTIS 1178
Db 2324 ALSGVLGVYDQAQNGALTLANGNRRYEGTITVAGELILGANGAFQGTSLNLIASGA-S 2382
QY 1179 GNTVSATVDTLTKSGSKIEAKSGE-----ANVT-----SATGTIGGTI 1218
Db 2383 ANINGYRQTVGAVNTGTVTLONGELTSTDTLNTLNGMLNVTGDLNLENGGASSISGGL 2442
```

```
QY 1219 SGNVTNVYANAGDLTV-----GNCAINATEGAATLTATGNTL---TTEAGSSITSTKQV 1271
Db 2443 TGNGI-LNTKGDFITSDNNGLAGQTNISDGSVTLGNGGTIGTGNLGSVDVLGDL 2501
QY 1272 DLLAONGSIA-----GSIN-AANVTLN-----TTGTLTVAGSDIKATSGT- 1311
Db 2502 NLVADN-SLANVISGDCGTINTTATVTLSONSFSGAHOICTNGELTVGQASNLGASSATV 2560
QY 1312 -----LVINAKDAK-----LNGDA-----SGDSTEYNAVNA-----SGSGSVT 1344
Db 2561 NLGTLTSHLILNGVSESIANVLGVAGSTVDIIGGADTALTANNSGFLGOYALAGNSKLT 2620
QY 1345 AAT-----SSSVNI---TGDLTNTVNCN-----IIKXDG-RNT 1373
Db 2621 VASTNNLGAASSVALAGTGTLSLGSFNGTSGNSVTGSGVLQVTDAAEVTLTSSNGVSA 2680
QY 1374 VRURGEIKEYIQGVASVEEVEIAKRVLEKVKDLSDERETLAKLGVSAREVEPNNT 1433
Db 2681 VTIDIADATLNL--DDIALFNHALTGNGLLNVAKNNDASTAFDFGATVGGATGTVNLNS 2738
QY 1434 ITVNTQNEETTRPSSQVLIISSEKACFSSGNGCARVCTNVADDG 1475
Db 2739 TFDLSGNTTVLAQATLKLSSGNLT-SVNGGVQNGITLAMNG 2779
```

## RESULT 10

```
Q9F285
ID Q9F285 PRELIMINARY; PRT: 3705 AA.
AC Q9F285;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE yaph protein.
GN YAPH.
OS Versinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277631; CAC14227.1; -.
DR HSSP; P22629; ISWG.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR01656; VACCYTOTOXIN.
SQ SEQUENCE 3705 AA; 370756 MW; 714FDF16455968C9 CRC64;
```

Query Match 8.0%; Score 594; DB 2; Length 3705;

Best Local Similarity 21.5%; Pred. No. 1.4e-11;

Matches 409; Conservative 262; Mismatches 695; Indels 538; Gaps 91;

```
QY 14 NALVAYSELA--RGCDHSTKESG---EKPMKVRHLAKPLSAMLISLGVTSIPQSVLA 68
Db 972 SGWTVTDLALNSGTVNIITGSGSDMDTDLATNVSTLEQDRAGSTLELINATVTDIDA 1031
QY 69 SGLQGMVDVHGTA-TMQVDGKNTIIRNSVDAILNMKQFNIDQNEVQFLOENNSAVENR 127
Db 1032 LDL----LVNGTAITSGTGQVSAIQGGSTVA-----NAHNYGLASSNSGDSGLVYN 1082
QY 128 VTSNOISOLGILDSNGQVFLINPNGITICKDAIN-----164
Db 1083 VT---LSALELLADGADALLATESGLTANR--VLNAELFGVGVLUVDAQNGALTANGS 1137
QY 165 NGCTASTLDTSNENIKARNFTFEQTK-----DKALAEIYNHGLITVKGDK 209
Db 1138 NRYEGTITVAGELILGANGAFQGTSLDILASGASANINGYSQTVGAVTVNGVTVLGSGG 1197
QY 210 SVNLIKKVKNEGVISVNGSGSLLAGQKITISDIINPTITYSTAAPENAEVNLGDIKAF 269
```

Db 1198 V--LTSGLTNGILDTGALNLTXXGASTVAGGLTGAGTINI-----N 1240  
QY 270 GGNINRAATIRNOGKLS-AD--SVSKDKSGNIVLSAKE--GEAEIGGVISAONOQAKG 324  
Db 1241 GGNLSVSAANSGLSGGTHADVASVLTDTGTGTSAVEVLGTNLNCAAAATN-----1295  
QY 325 KLMITGDKVTLTKGAVIDLSGRE--GGETYLGGDER--GEKG-----362  
Db 1296 --VLSGDG-TINTNAAVTLSSGNNSFGAHOIGTIGDTLTCQASNLGASSATVNLGLTSH 1352  
QY 363 ---NGI--QLAKKTSLEKSTINVS-----KEKGGFAIWMGDIALI-----DG 401  
Db 1353 LILNGVSESIANVLSGVAGSTVDIIGGADTALTANNNGEL--GOYALAGNSKLTVASTN 1409  
QY 402 NINAO-----GSGDIAGTGGFVETSGHD-----LFIKDAIVD-----AKEWLLDF 442  
Db 1410 NLGASSVALAGADTLLSLGFGNFTGNSVTGSGVLQVTDDBEVTLTSSNGVSNVAVTIDI 1469  
QY 443 DNVSINAED-PLN-----NTGIND-----EPTGTGEASDPKKNSLKTLLTNTT 487  
Db 1470 ADATLNDLIDALFNHVLGTNGLLNVAKNADASTAFDEGSTVGAFGINV-----LTNTT 1523  
QY 488 ISNVLKNAWTMTTASRKLTVNSINIGSNSHLLHLSKGGGCGVQIDG-----536  
Db 1524 FALSADNAAL-ARATLKLSDSVTTVGADP--TLHGLDLNGTLLFDGSPPOSQANGVV 1581  
QY 537 ---DITSKGNLTIY-SGGWVDVH---KNITL---DOGFNLITAAASVAFEGGNKARDA 585  
Db 1582 TVTDLANSGETISTGAGNENEHPTPPNVSLLEQDRGDLLELINAANVTGN-----A 1636  
QY 586 ANAKIVAGQVTIFG-----EGKDFRANV-----SLNGTG-----616  
Db 1637 NNLDLLVDGTAITSGTQGVESAIOQGGSTVANAHNYGLTSSNGGSGGLVYNYTLSALE 1696  
QY 617 ---KGLN--ITSSVNNLTNHL---SGTINISGNITINQTTKNTSYWQTSHDHSHWNSAL 668  
Db 1697 LLANGANALLATESGLTANRVLNAELFGVGLVVDQAQNGALTLANGNNRYEGTIVTAG 1756  
QY 669 NLETGANFTF---IKYISSNSKGLTTOYRSSAGVNF-----VNGNM 708  
Db 1757 ELILGANGAFQTSLLNLIASGASANINGYRQTVGAVTNGSGAVTLGNGVLTSGLLTNGGI 1816  
QY 709 -----SPNEKGAQVNF-----KLKPN-----ENMTSKPLPTREFLANITATG 746  
Db 1817 LDLTGGALNLAAGSSVAGGLTGAGTLNNGGDLAVSATNSGLSGQTHIADVASVLTG 1876  
QY 747 GGSVFEDIYANHSRGAEKLMSEINISGANFTLNSHVHVGDDAFKINKDLTINATSNFS 806  
Db 1877 TGTLL-----GTSAVEVLGTNL--NGAANMTNVLSSGGVINTNAATLSONNSFSG 1926  
QY 807 LRQ--TKDFFYDGYARN--AINSTVNTSILGCVNTLQONS-----SSITGN-ITTEKA 856  
Db 1927 AHQIGTDELAVGQASNLGASATVNLGTLTSHLILNGVSESIANVLSGVAGSTVDIIG 1986  
QY 857 ANVTLEANNAPNOQNRDVRVILKSLGVNLSLTGENADIKGNLTITSESATFKGKTRDT 916  
Db 1987 ADTALTANNNG-----FLGOYALAGNSKLTVASTN---NLGASSVALAG-AGDT 2032  
QY 917 LNTIG---NFTNNGTAEINITQGVVVKLGNVTNDGDLNIT--THAKRNQRIIGD-----966  
Db 2033 LSLSGFNXTFGNSVTG-----SGVLQ-----VTDDBEVTLTSSNVGNTVKVVDIADATLYVN 2084  
QY 967 -----IINKGSLNIT-----DSNDAEIQIGNISOKEGNLTITSSDKINIKOI 1011  
Db 2085 DIALDLHVLTEGTLNVAKYLATTAFAFXGTSVXXFSGIVNLTNTTFFALSADNAALARA 2144  
QY 1012 TTKKID-----GED-----SSDATSNANLTIKTKELKTE-DLSTSG 1049  
Db 2145 TLKLSDDSVTTGTDRILHGLDLNGGLIFDGPPOQANGVVIVTDLANSGETISITG 2204  
QY 1050 FNKAE-----ITAKDGRDLTI-----GNSNDGNSGAETVTFNNVKDSK 1089

Db 2205 AGNNEHPTPPNVNLSLEQDRGDLILLOLIDADNVNTGNANDELMINGTTSIAGQGVOST 2264  
QY 1090 ISADGNVNTLSKVTKTSSNGGR-----ESNSD-----NDTGLTITAKNVE 1130  
Db 2265 VQGGTIVANATHNYGMTSGCGSLYVNYTSLALELLADGANALLLATESGLT--ANRE 2321  
QY 1131 VNKDITSKTYNITASE-----KVTTTAGSTINATNGRASIITTKTGDISGT 1176  
Db 2322 LNAELSGVGLVVDRAQNGALTLANGNNRYEGTIVTAGELILGANGAFQTSLLNIASGA 2381  
QY 1177 ISGNTVSVATVLDLTKSGSKIEAKSGE-----ANVT-----SATGRIGG 1216  
Db 2382 --SANTNGYRQTVGAVNTGTVTGLNGGELTSTDTLINTGMINTGDLINLENGASSISG 2440  
QY 1217 TISGNTVNTVANAGDLTV-----NGABINATEGAATLTATGNTL--TTEAGSSITSTKG 1269  
Db 2441 GLTNGCI-LNKGDDFTISIDNNGLAGQTNISDQASVTGLNGGTIIGTNGSSVIDVLG 2499  
QY 1270 QVDDLAQNGSIA-----GSIN-AANVTLN-----TTGTLTVAGSDIKATPSG 1310  
Db 2500 DLNLVADN-SLANVISGDDGTINTTATVTLSGNSFSFGAHOIGTNGELTVGQASNLGASSA 2558  
QY 1311 T-----LVINAKDAK-----LNGDA-----SGDSTEYNAVNA-----SGSGS 1342  
Db 2559 TVNLGTLTSHLILNGVSESIANVLSGVAGSTVDIIGGADTALTANNNGFLGOYALAGNSK 2618  
QY 1343 VTAAP-----SSSVNI--TGDLNTVNGLN-----IISKDG-R 1371  
Db 2619 LIVASTNNLGASSVALAGTGTLSLGFNGTFGNSVTGSGVLQVTDDBEVTLTSSNGVS 2678  
QY 1372 NVVRLRGKEIEVKYIQPGVASVEEVEIAKRVLEKVDLSDEBETLAKLGSVAVRFEVN 1431  
Db 2679 NAVTIDIAATNL--DDIALENHALTGNGLNVAKNADASTAFDEGAVGVGAFTGVNLN 2736  
QY 1432 NTITVNTQNEFTTRPSQVLIISGKACPSGNGARVCTNVADDG 1475  
Db 2737 NSTFDLSGNTTIVLAQATLKLSSGNLT-SVGNGVQINIGTFLAMNG 2779

RESULT 11  
Q9ZHL0  
ID Q9ZHL0 PRELIMINARY; PRT: 4919 AA.  
AC Q9ZHL0;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Large supernatant protein 2.  
GN LSPA2.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=730;  
RN (1) TaxID=730;  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000.  
RX MEDLINE=990503326; PubMed=9811662;  
RA Ward C.K., Lumley S.K., Latimer J.L., Cope L.D., Hansen E.J.;  
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.";  
RL J. Bacteriol. 180:6013-6022(1998).  
DR EMBL; AF057696; AAC79761.1;  
DR InterPro; IPR003951; VerHae\_surfAg.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF03543; VerHae\_surfAg; 1.  
DR SMART; SM00235; ZnMc; 1.  
SQ SEQUENCE 4919 AA; 542595 MW; 5779201455CA69A0 CRC64;

Query Match 8.0%; Score 593.5; DB 2; Length 4919;  
Best Local Similarity 23.0%; Pred. No. 2e-11;  
Matches 385; Conservative 226; Mismatches 604; Indels 459; Gaps 83;  
QY 2 NKIYELFKSKRLNALVAVSELARGCDHST-----EKGSEKPARMKVR-----H 44

Db 3 NKRYKLIPSKVKNCVLPVAENIKSASGSSSSSKIAEDQEEPPDSLIACSLPLSSSIH 62  
 QY 45 LAL-----KPLSALLSLGLVTSIP-QSVLASGLQG-MDVHGTATWQV-----86  
 Db 63 LGLHNHSPKVPFGKNSVLLSL-----MPAQVWADSSNAIVDHSHGAKQFAVDERDPK 118  
 QY 87 -CNKTIIRNSVDALIN-----WKQFNI-----DQNEWVQFLOENN 120  
 Db 119 NGKEKVV-----VINIAKPDEOGISDNHSEFNIPNSAVFNNISKEGNSQLVGLLGENK 172  
 QY 121 N-----SAVENRVTSNOTISOLKGLDSGO--VELINPAGITIGKDALINTNGFTAST 171  
 Db 173 NLGSOAAKTFNQVTDQOESKISGLEVFGEKADLIINPVGVTINLVNGVKTINTDFEAST 232  
 QY 172 LDIENEMIKARNTEFTQDKALAEIVNHGLITVKGDG-----SVNLGGKVKNECVI 224  
 Db 233 SEVVEPHIKQLN-----VQRKVLIIGKGVATNGLSHDFVVAKNIBQOGKV 278  
 QY 225 SVNGGS-----ISLLAQKITISDIINPTITYSIAAPENEAVALNGDIFAKGGINVRA 277  
 Db 279 SIEGDSKPAKLANVFAAG-----NLTYD-----VNRDVI-----NRNT-- 312  
 QY 278 ATINQOKLSADUSKDSKNIVLSAKEGEABEIGGVISAQNOQAKGKLMITGDKVTLTK 337  
 Db 313 ----NPKKPTINTRKD---NIAIS---GE-----SAGSMYGRNIKFIIVT-----347  
 QY 338 GAVIDLSGEGGETVLGGDERGCKN--GIOLAK-----KTSLEK--GSTINVS 383  
 Db 348 -----DKGAGVNHQGVIFAEDDINILTDGNSRLNKVYADVVRVG 388  
 QY 384 KEGGFAIVWGDIALI-DGNINAQGGDIATKGGFVETSGHDLFIKDAINADAKEWLLDF 442  
 Db 389 K-----DIELANNGQIHADQOOLILNATGHVKLNDGSSVISNNLIGLSALN--LTL 436  
 QY 443 DNVSINAEDPLF---NNTGINDEFPPTGTCEASDPKKNSEL-----KTLT---TWTTI 488  
 Db 437 ENATVSANLNSFRVNTDKLNNLSKVSARAADLOSGLNLDKASVLAHKLFLINISNDVSL 496  
 QY 489 SNVLK-NAWTMNITASKRLVNS-----INIGSNHLLHLSKQORG-----V 532  
 Db 497 NNGSKUSANLKIKKVRDLNLSSELSANLTLNNTITLKNKSKFTAGNLTNNTVNV 556  
 QY 533 QIDGDTISGKNLTIYSGGVVHVKNITLDGOF-----LNTI-----AAS 572  
 Db 557 TLNDSLEAANNLTL-----NWKVNLNDASKLSANKLIDLNVDTNLTLSKTSLSAGE 610  
 QY 573 VAFEG-----GNKARDAANAKIVAQGTVTITGEGK-----DFRANVSLNG-----TG 616  
 Db 611 LTFKVKVNTLNDSLEAANNLSNASHNVTLNKSKLSAQKADIKAVNLTLNDLTETA 670  
 QY 617 KGLNISSVNLTLNLSGTI-NISGNITINQITRK-----NTSYWOTSHDSHW-WV 665  
 Db 671 KNLDI-----NSTITNGTITAGIFANITTEKLNKKEKALLAEQNLNF--TVNGSHYENK 724  
 QY 666 SALNLETGANFTFK---YISSNSKGLTQYRSAGVNFNGVN-----GNMSFNKE 714  
 Db 725 GDVSKDQATVTSKNSDTSNGSLVNAQNLKVNYNFTISQGGDITLIGNVTNLN-AS 783  
 QY 715 GAKVFEKLPKNENMNTSKPLTRFLANITATGGGVFFDIYANHSGR-----GAEKWSI 770  
 Db 784 GTFN-----SGNLTIVKLDVGDQNTFNKGNLTIVGEDLHLSKTKINDGKLISIKNL 838  
 QY 771 NISNGANTLNSHVGDGDAFKI-----NKDLTINATNS--NFSLRQTKDDFYDGYARN 821  
 Db 839 NTSSEADFINNGTLIGTEALKIATKGNFTNKEKAILASLLDISVAEGKKTENN-----894  
 QY 822 ANSYNINISILGGVNTLGGSSSSITGNITIEKAANYVLEANNAPQOINRVRVVKLS 881  
 Db 895 TIESCKNLNITNGAFNLVDNATIRSGVLNITSGNV-----SNN-----GT 937  
 QY 882 LLVNGSLGTGENADIKGNLTISESATFKGTRDITLNTITNFTN--NGTAEIN-ITQGVV 938  
 Db 938 LISNERLNTS-----AANFTNESNGTVMNSGLNLIIA 970

QY 939 KLGNTVNDG-----DLNITTHAKR--NORSIIGG-DIINKKGLSLNITDSN---NDABIQ 986  
 Db 971 KOGNITKNLIASROQLNLTAVADNITNDSNISKIIVLHSLGNSLKQDQVYNLGEIY 1030  
 QY 987 IGGNISOKEGNLTISSDKINITKQITLKK-----GIDGEDSSSDATSN 1031  
 Db 1031 AGNISVKAHL-----KNVDKLMGDITTKTKEGQASYKLYQASNGHFGNDGSS-GYSBGD 1086  
 QY 1032 LTIKTELKLTEDLSISGFNKAETARKDRDLTIGNSNDGNGSGAEAKTVTFNNVSKIS 1091  
 Db 1087 LNIKGFADLNDKLTQVRIQR--IYA--GRLTFNKSNAAGK-----SEII 1128  
 QY 1092 ADGHNTVLSKPKTSSSNGGREGSNDNDGLTIFAKNVENVKDOIITSLKTVNITASEKVT 1151  
 Db 1129 NRG---TIIVKNKLSYSDSVFNNMQSOKVDLYTKIFEAKSDIE--LTFKTNCTHPVYL 1183  
 QY 1152 TAGSTINATNGKASITTKTGDISGTISGNTVSATVDLTT-KSGSKIEAKSGEANVTS 1210  
 Db 1184 NFKSNNEKKYRNSENTKNFKSIGDLNEALSASAPAEIAYYSGS-----SSNINPVSY 1239  
 QY 1211 TGTIGTISGNTVNTANAGDLTVNG--AEINATEGAANTLTATGNTLTTPAGSSITSTK 1268  
 Db 1240 LAALGNANSSPHYLNTALKHIGNGWODDLKQENIKVKKOWEDEFKKDKGAS-----1294  
 QY 1269 GOVDLLAONGSTAGSINAANVTLTGTLTTVAGSDIKATSGTLVINAKDAKLN-GDASG 1327  
 Db 1295 KMLDLYPNTDKAKI-FAGIIRNGNDTISDVESEDEFKKYSKFQ-NGEAKNDGTGDSY 1352  
 QY 1328 DSEVN-----AVNAGSGSVTAATSSSVNITGDLNTVNGNLNISKDGRNVRURGK--E 1380  
 Db 1353 DSTKASEKVKVENVDHENI---DBHKLNGKHETITPGVSPENLNKNDHQDKLGE 1409  
 QY 1381 IEVKYIQGVASVEEVTEAKRVLEKYD--LSDEERETLAKGVSAVRFPVPPN 1432  
 Db 1410 IDKSIISSELLA---QPVYTEKSAARSDPRVQNDKEALDNLRYTRLVSYNQNN 1460

RESULT 12  
 Q8XP07

ID Q8XP07 PRELIMINARY; PRT; 2737 AA.  
 AC Q8XP07;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Probable hemagglutinin-related protein.  
 GN RSP1539 OR RS05770  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salenoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646085; CAD18690.1; -;  
 DR InterPro; IPR000566; Lipoclin\_CyLFABP.  
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 2737 AA; 268490 MW; 12F383C02C101D0C CRC64;

Query Match 7.68; Score 561; DB 16; Length 2737;  
 Best Local Similarity 22.4%; Pred. No. 1.2e-10;  
 Matches 326; Conservative 187; Mismatches 546; Indels 396; Gaps 64;

QY 82 TMOVDGKNTIIRNSVDAILNWKOFNIDONEMVQFLOENNSAVENRVTSTNOISOLKG--- 138  
DB 155 SLTVDGVGLIINNSLAGGGTFLGNGVGN--ANLAASGPASTILTQVGTGPEIRINGTVE 212  
QY 139 ILDSNGQVFLNPNPITIGKDAIINTNGFTASTLIDISNENIKARNFTFEOTKOKALAEIV 198  
DB 213 VFGSPASVIFAAPAGVYTOGAGFTNTPRVTLTSTGTPQFLNGSGANVAFDQA--TAVGFLV 270  
QY 199 NHGLTIV-----KQDGSVNLIGKVKNEGVISVNGSGISLLAGQKTIIDII 245  
DB 271 NSGRIEIPAAGSTAGAGIEGTVGAINLIGTV-----GVNAPLEAGNQINV--IA 319  
QY 246 NPTIYSTAAPENEAVALGDIFAKGGINVRAATIRN-----QKGL--SA 288  
DB 320 GNOQVAPVATGTGRAGSDQWVSGAGANAANSASQAQNGLAIDATAFGAMTAGQIKLISTA 379  
QY 289 DSVSKDKSNIVLSAKEGEAEIGGVISAQ--OQAGGKLMITGDKVTLKTGAV---ID 342  
DB 380 QGLGYRAAGDLAANTSNVNDANGDVSGVNYGQTAG--ITTTGSVST--SGAVRAQOD 435  
QY 343 LSGKEGGETYLGDERGEGKNGIQAKKTSLEKSTINVSKEKG--GFAIWMGDIALID 400  
DB 436 VTIGAGGDDVTLGG-----AAQAGNVTVSAGNVAGSGDLAAVHGLSVSAGNSANLG 487  
QY 401 GNINA-----QGSQDI-----AKTGGFVETSCHDLFIKDNAIVDAKEMLLDPDN 444  
DB 488 GNLNAASIAVTAQGGKDGTDITLGGKVASPGGIALNAARDTTIAGQ-----LTGSG 539  
QY 445 VSIANAEDPLFNNTGINDFPTGTGEASDPKKNSELKTLTNTTISNLYLNKAWTMTNITASR 504  
DB 540 VSVNAGRNLVSGAIG-----SVGDNLNLAAGASVSTTGAVTT-----QANLTASA 585  
QY 505 KIVTNSSTNIGNSHLILHSGKQGGVQIDGDIITSKGNLTIYSGGVVDVHKHITLDQ-- 563  
DB 586 QDVRLLGGTAAHGHVAIQANA-----GSITT--AGLTAAODIALTAGQONATLGAA 635  
QY 564 ----GFLNITAAVAFEGGNKKAADAANAIVAOQTVTITGEGKDFRANNVSLNGTKGL 619  
DB 636 TOTGNGLTVTAGNTVGGAGNLASKAID--VQAGGSVDVSG--NVSANRIAMQAAGR-- 688  
QY 620 NIISVNNLTHNLSTINISGNITINOTTRKNTSYWQISHDSHNVVSALNLETGANETFI 679  
DB 689 ----DGVGDI--RLGGNVGAPGTITLN--AARDTTIAGSVSDSLN----- 727  
QY 680 KYISSNSKGLTTOYRSSAGVNFNGVNGMSFNLKEGAKVNFKLKPNENNTSKPLPIRFL 739  
DB 728 -----LATQRLNSVGVVGVSTKGNVSLTARTGA----- 755  
QY 740 ANITATGGGVFFDIYANHSRGAEELKMEINISGANFTLNHVRGDDAFKINKDLIN 799  
DB 756 --VTT-----QGAVTTPGNLVSSCADTSLGGQVSAAGTASVN----- 791  
QY 800 ATNSNFSLRQTKDDFYDYARNAINSTYNISILGONVTLGGQSSSSITGNITIEKAANV 859  
DB 792 -----AGGNLSTAGQISGN--GTLTLNAGONL 816  
QY 860 TLEANNAPNQNIIRDVIRKIGLS--LLVNGSLSTGENADTKGNLTISESATERGKTRDTLN 918  
DB 817 TIGGQVGS-----ADANLQAGSNVTVNGALTSTG--NASVLAGQSGIA----- 857  
QY 919 ITGNETNNGTAEINITQGVVVLGNVTNDGDLNITTHAKRNQRSI--IGGDIINKKGLNIT 977  
DB 858 LAGDVAAGGNATLDAQ-----TITGPNLSAAQATAKVTGSDIDLGGQVKQKQVALTAN 911  
QY 978 DSNDAEIQGNISQEGNLTISSDKINITKOITI--KGIIDGEO-----SS 1023  
DB 912 GSNLGDVRLGGVAG--APGSVTLSA-----TRDATLGSNAIAGGDLTATAGRNLVNGAA 965  
QY 1024 SDATSNANLTIKTELKLEDLISGFNKAETAKDGRDLITGNS--NDGNSGAPAK---- 1078  
DB 966 ASVNGNVNLTAAQOALASTGSIQA-----NOGDVNTATAGQGLNVGSGSVYAGRNAGLVAQGN 1022

QY 1079 -TVTFNNVKDSKIS-ADGHNVTLNLSKVKT-----SSSN-----GGR----- 1112  
DB 1023 ATVSGNLTSLGKASISGGNTTSLGQKTLGGDQLQASAAALSVQALNVYVGNAILRGTDI 1082  
QY 1113 -----ENSNDNT-----GLTTAKNVEVKN-----DITSLKTVNITASEKVTTPAG 1154  
DB 1083 AVGSAAGOSNAVQGTLDVAVASRGLTL-AGNSNANALNGLGATIVNOGSTLATQRTATVSG 1141  
QY 1155 STINATNGKAS-ITTKTGDII--SGTISGNTVSATVDLITTKSGSIEAKSGEANVTSAT 1211  
DB 1142 MVTNAGMLAANQLTVSATDLVNRGTVGGQAVKLNTTGNLDNAGGLVVGSTLDVVTAGALT 1201  
QY 1212 GTIGGTISGN-----TVNVTANA-----GDLTVNGGAEIN-----ATE 1244  
DB 1202 SNRGGTFFGDLAGKSPPTTGNLTFTVNGGAGSFNNAGGOLLAGNLTNTPNQAFDPSAA 1261  
QY 1245 GAATLTATGNTLTTEAGSSITSTKGQVDDLLAQNGSIAAGSINAA-----NVTI 1291  
DB 1262 TACTLNA-NNTLT-----SIOSINNTGTWNVQSSVA--INAAQGFISGCTIOKAGNLSL 1314  
QY 1292 NITGTLT-----VAGSDIKATGTLVINAKDAKLDASGDSTEVNAVNASGSGSVTAAT 1347  
DB 1315 STAGALANSOIVGSGSNVALSAGTLT-NTGTHADGNLALAGNVNAGTAEALGNI-AVT 1372  
QY 1348 SSVNITGDLNTVNG 1362  
DB 1373 GSNYDNOGGKTOANG 1387

RESULT 13  
Q9HWU6 PRELIMINARY; PRT; 1018 AA.  
ID Q9HWU6  
AC Q9HWU6  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
DE Probable adhesin.  
GN PA4082.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
CX NCBI\_TaxID=287;  
RN [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Keizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RL Nature 406:959-964 (2000).  
DR EMBL; AE004824; AAG07469.1; -.  
KW Complete proteome.  
SQ SEQUENCE 1018 AA; 100433 MW; F373734D77FFA94D CRC64;

Query Match 7.5%; Score 553.5; DB 16; Length 1018;  
Best Local Similarity 22.6%; Pred. No. 7.4e-11;  
Matches 301; Conservative 181; Mismatches 453; Indels 399; Gaps 61;

QY 1 MNKYLKFKSRNLNVALVAVSELARGCHDSTKESG--KPARMKVRHLALPLSLALLSL 57  
DB 1 MNKVAL-----VNVVSGQWNVVSESGRRRRKPKAGAKA-----AIAVLALL 43  
QY 58 GWTST-POSVLASGLQGMVDVVGHTATMOV---DGNKTIIRNSVDAIL-NWQFIDQDEM 112  
DB 44 CATALAPAYALPS---GGTVVGGSSANGEIHLSSGNSLVNOKVKLIANWDSFVAAGER 100  
QY 113 VOFLQENNSAVNRVTSNQLSOLKILDSNGQVFLNPNGITTKDAINTNGFTASTL 172



Db 101 VIFNPSSSIALNRVIGTKASDIQGRIDANGQVFLNPNVGLFGRGAQVNVGGLVASTL 160  
QY 173 DIS-----NENIKARNFTFEQTKALAEIVNH-GLITVKGDSYVNLGGKVKNEGVSIVN 227  
Db 161 DITDAEENGSSRYRFTGPTNG-----VLNHGGAITAAEGCSIALGAQVDNRGTVLAQ 215  
QY 228 GGSISLLAQKITISDIINPTITVIAAPENEA- -NLGDIKFAKGGNINVRAAT----- 279  
Db 216 MGGVGLGAGSDLTINFQGNLLDIRVDAGVANALASNGLLKADGGRVLMARATANALLN 275  
QY 280 --IRNOCKLSADSVSKSGNIVL- SAKGEFAETGGVISAQ--NQAKGKGLMITGDKVT 334  
Db 276 TVVNSQAIARSLL-RGKNGRIVLDGPGDKVMVGAGSALNPNPGHGTVEVRGAQE 334  
QY 335 LKTCVAVIDLKSGEG--GETYLGDERGEGKNGIQLAKRTSLEKSTINVSKEKKGFAIV 392  
Db 335 VALGTQVNTLASNGLNCTWKAIDKID-----VRPSAVSDGVTVHADTLRSN----- 381  
QY 393 WGDIALIDGNIN-AQGSDDIAKTGGFVETSGHDLFIKDNAIDVAKEWLLDFDNVINAED 451  
Db 382 -----LASTNIELVSTKGDLDLGSVNVASGNRL----- 410  
QY 452 PLFNNTGIDEFTGTGEASDPKKNSEKLTTLTNTTISNLYKNATWNTITASRKLTVNSS 511  
Db 411 -----GLGSAADLTNLNGLNAS-----GAKAGLELKAEGA 440  
QY 512 INTGNSHLLHSKG-----ORGGGVOIDG--DITSKGGMLTYSGWGDVHKNTLDQ 564  
Db 441 IDI--NDKIVLGGAGSALAMDAGEHRVNGTASVSLAGANATYVSGGY-----YTQVN 493  
QY 565 FLNITAAVAFEG-----GNKARDAANAKIVAOGTITTEGKDFRANVSLNGTGKLN 620  
Db 494 LAQLQAINKLDGLYLGN-----ILGSIYCT-----ALQSIGGPAG 532  
QY 621 IISVNNLPHNLGNTINISGNITFINOTTRKNTSYWTSKSHDNVNSALNLETGANFTFIK 680  
Db 533 VFSG-----TLDLGNSIGNLSIS-----NTGPNV----- 557  
QY 681 YISSNSKGLTQVRSAGVNFVGNVNMSEFNLEKAGVNEFKLPNENMNTSKPLPRLFLA 740  
Db 558 -----GLFA--RSS-----GTLNKLN-----NLVSDNTYTGSPSSIGALV 593  
QY 741 NITATGGSVFFDIYANHSRGAELKMEISNIGANFTLNHVRGDDAFKINKDLTINA 800  
Db 594 GI-----NSGRIANVSAGSVV--VGSRLSNALGG--LVGRNISGGI 632  
QY 801 TNSNFSRLTKDQFDYGYARNAINSTYNSILGGVNTLGGQNSSSSITNITIEKA--ANV 859  
Db 633 ANASVSGGVNT-----GYAAS-----TAVGG--LVGENFTTAGPEAVIENAHNSV 675  
QY 860 TLEANNAPQONIRDRVILKGLLVNGLSLTGGENADIKGNLTISBSATFKGT---RDT 916  
Db 676 HVAQAOSTERNS-----LGGVGLVGLNA--KGMIRASGS---QKQVETIRPG 717  
QY 917 LMITG--NFTNNGTABINTQGVVKGNTVNDGDLNITTHAKRNQRSIIIGDIIKKGSL 974  
Db 718 LVNGLVGVNMFCHVSDSSASQVAGAGNAGTGLVGLSSGGEIFRSQASGSVYSGG-- 775  
QY 975 NITDSNNDAAEIOGGNISQEGNLTISDPKI--NTKQITIKKIDGEDSSSDATSNAML 1032  
Db 776 -----LATGGLIGKAEKNGMLGNLUGKASGVTDQ-----GGADLGLLVGNNSOS 818  
QY 1033 TIKTKELKLTEDLSISGFNKAEITAKDRDLTIGNSDGNCSGAEAKTFTFNNVKDSKISA 1092  
Db 819 ALETAE---ATGKVSAGGSNRVGGGLIGNL-----GGSAHAIRSGDV-----S 859  
QY 1093 DGHNNTLSKVKVTSNNGGRESNNDNGLTTAKNVEVKNKDTISLKVNTITASEKVTTT 1152  
Db 860 GGFNSLVGLV---GHNGG-----ELNVDPASGRVSA 889  
QY 1153 AGSTINATNGKASITTKTKDIDISGTISGNTVSVATVDLTTKSGSKIEAKSGEANTVSATG 1212  
Db 890 ASASV-----GGLVGSNAGSILSARSSTVNGSGRSRIGLVLGNQI---OG 933

QY 1213 TIGGTISGNTVNTANAGDLTVNGCAEINATEGAATLTATGNTLTERAGS-SITSTKGOV 1271  
Db 934 RIVSMSSBGTV-----SGDYVYSMG-----GLAGL-----NLGSIEYSGVSGKI 972  
QY 1272 DLLAQN--GSIAGS 1283  
Db 973 DFKPQSHYQIYGA 986  
RESULT 14  
Q98E20  
ID Q98E20 PRELIMINARY; PRT; 3930 AA.  
AC Q98E20;  
DT 01-OCT-2001 (Tremblrel. 18, Created)  
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Hypothetical protein ml14444.  
GN ML14444.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti";  
RL DNA Res 7:331-338(2000).  
DR EMBL: AP003004; BAB51100.1; --  
DR InterPro: IPR001425; Bac\_rhodopsin.  
DR InterPro: IPR000734; Lipase.  
DR InterPro: IPR003880; Pplantne\_attach.  
DR PROSITE: PS00327; BACTERIAL\_OPSIN\_RET; UNKNOWN\_1.  
DR PROSITE: PS00120; LIPASE\_SER; UNKNOWN\_1.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 3930 AA; 380662 MW; 9ACD1ACA185BF712 CRC64;

## Query Match

Best Local Similarity 21.9%; Score 548; DB 16; Length 3930;

Matches 366; Conservative 227; Mismatches 641; Indels 441; Gaps 82;

QY 56 SLGVTSIPOSLASGLQMDVHGTYATMQVDGNKTTIRNSVDALINMKQFNIDQEMVQF 115  
Db 861 SLGGVPI-----NITGDNVAHDVAISDPGGAATLTNSGSAVVT-----AANGNSLLDF 909  
QY 116 LOENNNSAVENRVTSNQISOLKILDNS-----GQVELINPN-----GI-TI 156  
Db 910 NLSGGSDAGINATISINGLI- IQGVTYVNDVTGLFSGVTGTVSDNLNVONASQTGIELL 968  
QY 157 GKDAINTNGFTASTLDISNENIKARNFTFEQTKALAEIVNHGLITTVKGDGSVNLIGG 216  
Db 969 GSSATVFTG-NATIGNATNVGLSANFDDGTATFD-----DLDTGG 1009  
QY 217 KVKNEGVISVNG-----GSISLLAQKITISDIIN--PTITYSIAAPENEAVALGDIF 267  
Db 1010 GI---GVGLVNGSSGTLTFGAGSSIAGTSNAPFSINSTPNVNTYNGTISQTSAAVAKIS 1066  
QY 268 -AKGNINVRAATIRNOGKLSADSVSKSGNI-VLSAKEGEAEIGGVISAQNOQAKGK 325  
Db 1067 NMTGGTATFGMTASTGAANALDLSNAGGAINFTGGDLDTADGFAFSA-----VGGGT 1122  
QY 326 LMITGDKVTLTKGAVIDLKSGEGGETYLGDERGEGKNGIQLAKRTSLEKSTINVSXKE 385  
Db 1123 LRVTSNNTVATG-----LGPDEK-----GQIA-----GMTIGAAG-- 1154

QY 386 KGGFAIVMGDIALIDGNINAQSG-DIAKT-GGFVETSGHDLF-IKDNAIVDAKWLDDP 442  
Db 1155 -----VNESSVA-VDSGGGVSTGAIASNTSGNVTEGGVDLFRISGTAL-----SL 1200  
QY 443 DNVSLNAREDPFNNTGINDPEPTGTGEASDPKKNSELKLTITLTNTTISNYLNKAWTMMITA 502  
Db 1201 DNASGCIYS--FNGTTRKVDTV-VGGPG-----FLVNSAAATVGNLVTNNI-- 1245  
QY 503 SRKLTVNSINIGNSHLILHSKQORGGOVIDDITSGK-GNLTIIYSGGWDVHKHNTL 561  
Db 1246 -----AGADVSLTNN-----GTIGIGTIIINSNGTGVVVSAGSAAI--TVSA 1287  
QY 562 DOGLFNITAAVAFEGGNKNKARDAANAKIVAQGTVTITGEGKDFRANVNSINGTGKGLNI 621  
Db 1288 D-----ISSSATAPGTAVKVDGTGGSVTFSGLTSTGTGTGVSVSNTAA-GSGVGFGA 1340  
QY 622 ISSVNLNTHLSGTINISGNITINQTR-----KNTSYWQTSKSHDHSNWSALNLET 672  
Db 1341 V-TVSGAAGNIGISGNAGSVTFGTNTVTLGSAANAAGINFSCTNADVLFGTNTVTGA 1399  
QY 673 GANFEIKYISSNSK--GLT-----TOYRSSAGVNEGVNGVNMSENLKEGAKVNFKLKP 725  
Db 1400 GANQTGIDFSGSSATAGGLTITGTGLTSRGIDLSSTTGKNTITFARGSSIT----- 1453  
QY 726 ENMTSKPLPIRFLANITATCGSVFFDIYANHS-GRGAELKMEISINSGANFTLNHV 784  
Db 1454 -NVG-----VGVLSGGTATYASANANFTFGD-----NAGDGLQSTISAAA 1494  
QY 785 RGDDAFKINKDLTINATNSNFSRLQTDFFDGYAR--NAINSTYNISILGGNVTLGGQN 842  
Db 1495 GGYVTNTIGLPTLG--NYNFN-----DVFTGSAHLASAVGGTIVMSQMGVIAAGTDG 1547  
QY 843 SSSSITGNITIEKAANV-----TLEANNAPNOQNRIDRVIKLG--SLVANGSL- 888  
Db 1548 LSAAVT-TISVAQDALAGTAGTFAFGVGTVDLSSTPFTLDSQOSTGTGNNNTSIVSGTIQ 1606  
QY 889 -----SLTGNADIKG-----NLTISEATEFKGTROTTLNT--TGN 922  
Db 1607 PVNVOGNLGSAGGNVTNEGVTGTGDFPLHLLGSNQVRNTAFNFSAGSGSVFNVDSAGG 1666  
QY 923 FTNNGTAEI-----NIFQGVVVLGNVTNDGDLNITTHAKENORSIIGGDIINKKSL-N 975  
Db 1667 FNNAGGIVIEGTVSVNATGTAFAKVLGDSNLSIT-----NNNVNAGTLLDVNGGTGN 1721  
QY 976 IT-----DSNDAEIOIG-NISOREGNLTISDKINIKOITIKKIDGEDSSSDAT 1027  
Db 1722 ITLRGTLPNATPGLTGGISIANRSGGL-----VNFTDKVIV-----GGAGVSLT 1769  
QY 1028 SNANLTIKTKELKLTEDUSIGFNKAETAKDGRDLTIGNSDNGNSGAEK--TFTFNN 1084  
Db 1770 GNTGTVTFADL-----DITTSGATAFASGGGTVNVTTGTINATNAQAAALDGTAGINF 1825  
QY 1085 VKDSKISADGHNVTLNSKVKTSSNGGRESNDNT-----GLTIT 1125  
Db 1826 ASTSATFASGNGIDQLNLSGTFSGTGLTNTGSGTGSFNVGSATNLSGNAVIVSGGTIA 1885  
QY 1126 AKNVEVNRDITSLKTNITASEKVT--TTAGSTINAT--NGKASITTKTG----- 1171  
Db 1886 SNGTGAASVIOELTGGSTVLSGNVTDGLAGAGGNIVVTGIDNGTAATVTFSGSKQIDSG 1945  
QY 1172 -----DISGTTISGNVSVSAVDLTTKSGKIEAKSGEANYTSATGTI-----GG----- 1216  
Db 1946 ATDGVSLLGNPGLTAFISGLVITTSAGAGFHAS-----TFTGTIVTVTKDGAANNI 1999  
QY 1217 -TISANTV--NVNAGADLT-----VNGAEINATEGAAT-----LTATGNT 1255  
Db 2000 TTTTGNALNDNVTVGVGGINFDSISSNGTGTGIALNVVSGGANLGAVDLEGITSRGD 2059  
QY 1256 LTFEAGSSITSTKGVDDLLAANGSTAGSINAA-----NVTLN-----TTGLTITVAGSDIK 1306  
Db 2060 VSGTLGSLNFTSLNIGLNAAN-AIGLDLSAALGVSNITAGDFDVGGSFAGTIGIDMA 2118  
QY 1307 ATSGTLVINAKDAKLNGDASGDSTEVN-----AVNAS--GSGSVTAATS----- 1348

Db 2119 GTTGTCTIOLGDTVNNNPAGQSTIANVGYGVQFSSATNAQLVFGDGAGPAESSIATTTGG 2178  
QY 1349 -----SSNYITGDLNTVNLNII-----SKDG--RNTVLRGKEIB- 1382  
Db 2179 OVTHATDITPTSGDYFNDFNSGDTSNLSAVRVYVTVAGTGDGLANPGSYLGAQAST 2238  
QY 1383 ---VKYIQGVASVEVIEAKRVLEKVD-----LSDEERTLAKGV-----SAV 1425  
Db 2239 ANVVVLIDKNVGAQITDLSGTTFLNDDGQVLLAFKSGDAADVDSQVDTSGGSAAF 2298  
QY 1426 REVEPNNTITVNTQNEF--TTRP-----SSOVI--ISEGKACFSSGNGARVCTNV 1471  
Db 2299 HFTTQNTPLISAPGGIDTLRPVLOSNNATSVINFATSGTGTGTGIIENLIVSNV 2353  
RESULT 15  
Q8RDQ9 PRELIMINARY; PRT; 3165 AA.  
AC Q8RDQ9;  
DT 01-JUN-2002 (TREMELREL 21, Created)  
DT 01-JUN-2002 (TREMELREL 21, Last sequence update)  
DT 01-JUN-2002 (TREMELREL 21, Last annotation update)  
DE Fusobacterium outer membrane protein family.  
GN FN1449.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteriia; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
RA Larsen M., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fonstein M., Kyripides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010650; AAL95642.1; -;  
KW Complete proteome.  
SQ SEQUENCE 3165 AA; 331849 MW; 6DFE562C8490DEAL CRC64;  
Query Match 7.3%; Score 543; DB 16; Length 3165;  
Best Local Similarity 22.5%; Pred. No. 5.6e-10;  
Matches 372; Conservative 222; Mismatches 535; Indels 524; Gaps 85;  
QY 68 ASGLOGMDVVHGTATMOVDGNKTTII-----RNSVDAIINMKQFNIDQNMVQFLOENNSA 123  
Db 396 AVGLYSADPTKFNNTFKIESGKTLDELKGNSTFGLLGNNT-TVTNPSLLSKYLNNTSD 454  
QY 124 VFNRTVSNQ-----ISQLGILD-----SNGQVF-----LINPNG-----ITIGKDAIN 163  
Db 455 KINIVSFGEGASLFYATSKAKAILDEYKVTNGDAISTAVLVANNGANVEIASGKKLETN 514  
QY 164 TN-GFTASTIDISNENIKARNFTFEQTK-DKAL-----AEIVNHGLITVKGDSVNLIG 215  
Db 515 TNAGLIAINGTVGFTSVAKNNGTLLSTRIDKGIYVTSAAANGENSCTITMONKNAVGLIG 574  
QY 216 GK--VKNEGVISVNGSGISLLAGKITTIDINPITTSIAAPEANVLNGLIFAKGN 272  
Db 575 SKGSLNKTCKIELEAVS---SAGVYAEADSNMTNSGTTSEIIV--NKEASVG-IYAKEYS 628  
QY 273 INVRAATIRNOGLKLSADSVSKDSGNIVLSAKEGEAFI-----GGVISAQNOQAKGKLM 327  
Db 629 ISSVKNVKNREGKIEKADGDGKSAGI-YSKKEGKALTIENTGTNEVQAQKASA---GIY 684  
QY 328 ITGDKVTLTKGAVDILSG--KEGETYLG--GDE-----RGEKNGIQIAKTSLEKSGT 378  
Db 685 TKNESTQANTQSEVTNSGLVYKMSAENSIGIMGEKSKITNTGTGTGKIEIVEK----- 736  
QY 379 INVSGKEKGGFAIVWGDIALIDGNINAQSGSDIAKTGGFVETSGHDLFIKDNAIVDAKEW 438

Db 737 -----KSAGILYNESAVIN-----SGRISLNSISSSSDGL----- 769  
Qy 439 LLDENVSINAEPLFNNTGIDNEPPTGTGEASDPKKNSELKTTTLTNTTISNLYKNWATM 498  
Db 770 -----VGISVDG-----SSTGEND----- 783  
Qy 499 NITASRKLTIVNSINIG-----SNSHLLHSGQORGGVQIDGDTSGGNNLTIVSG 550  
Db 784 ---ASGEIKVDAAYSTGMLSSGGDVNAGKIALEKESGVMYATNANVTNSG---AIPKG 837  
Qy 551 GWVDHKNITLDQGLNITAAVASAFEGGNNKARDAANAKIIVAQGTVTITGEGKDFRANV 610  
Db 838 IFIKDEKSVGI---YSKINSSI-----ANKVTNSGTIDTAGTSKTSAGIY 882  
Qy 611 SL--NGTGKGLNISSVNNLTHNLSGTINISGNTIINOTTRKNTSYMQTSHDSHWNVSA- 667  
Db 883 SIIENGATKTL---SAVNN-----GNITING--KKSUVIYAKNESSHANTESD 925  
Qy 668 -----LNLETGANFTFIKY---ISSNSKGLTQYRSSAGVNFVNGVNMSEFNLEK 715  
Db 926 VTNSGKIEVKNESAGVLAEKYKVTNGSGTNGIIVSAKSA-----GIICKLGSEI--- 977  
Qy 716 AKVNFKLKPNB-----NMNTSKPLPIRFLANITATGGGVFFDIYANHS-- 759  
Db 978 --INSGLIKTETATPTVATDGVVGLNSNK-----ATNTSGGVITLD--TNYSTG 1024  
Qy 760 --GRGAELKSEINISGANFTLNSHVGGDAFKIND-LTIN-----ATNSNFSLR 808  
Db 1025 MYGEANSQLTNEGNT-CTNKEYIVGMAGDSSTVTNKNITLNGKATGIFGKNSSTLN 1083  
Qy 809 QT-----KDDFYDGYARNAINSTYNISILGGNVTLGG----- 840  
Db 1084 ETTCKIPTKEESVCMYSSSLKATNGITITEKTSAGMLGDKANIENDSSIPTKEEMS 1143  
Qy 841 -----QNSSSSIT--GNITIEK--AANVTLEANNAPQOINIRDRVIKLGLSLLVNGSLSLT 891  
Db 1144 AGMVKNGTSKATNKGTVTTEKTSAGILAEIDEA-----NGG-TVS 1184  
Qy 892 GENADIKGNLTISP--SATFGKTRDTLNTGNTNNGTAEBINITQGVVKGNTVNDGDL 949  
Db 1185 GLN-ETTGTITVSETSAGMLGVKSA-----VTASTAKL-SLTNKKDI 1226  
Qy 950 NITTHAKR-----NORSIIIGD--IINKGSLNITDSNDAEIOIG-----G 989  
Db 1227 NINTKNSAGIMVYNVESTAVKENVLAENTGTINLTSSATNEKNIGILANKATGINTGNI 1286  
Qy 990 NISOKE--GNLTISSDKINITKQITI--KKGID--GEDSSSDATSNANLTIKTEL--KL 1041  
Db 1287 NINSKESIGMLGQNASSTNNKTITLSGEGIGMLSKDTSIADNDNDIINVNGKESLGML 1346  
Qy 1042 TEDLSISGFNKA-EITAKDG-----RDLTIGN-----SND 1070  
Db 1347 GEDSGTVKNKNTISVTAERGIVFVRDNGVKGSGTGENTSTGTITILENKEAVGIFAKNN 1406  
Qy 1071 GNSGAEAKTVTEN-----NVKDSKISADGHNVTLNSK-- 1102  
Db 1407 GTSDSAKNSGTINLKGADGSTIKESLIGMPAQAEAGKANKVNTK-----DINVTKKS 1460  
Qy 1103 -----VKTSSN-----GGRESNDNDTGL-----TITAKNVEVKNKDITSL 1138  
Db 1461 VGIYAKNDASNITDVLNTGDNININSKESAGVYAPKANISKVGTITLKN---SIDSNGS 1517  
Qy 1139 KTVNITASEKVTTTAGSTINA-----TNGKASITTKTDISGTISGNTVSV--- 1184  
Db 1518 SAVVVSCKGKVANTAGVKINLGTVQNVRVAYVYVNGKDS--ALAGADIGKITGYGVGYLQ 1575  
Qy 1185 -----SATVDLTTKSGSKIE-----AKSEANVTSATGTIGTISGNTVNT 1226  
Db 1576 GTSGDKATLD---KNTSKLDYTLQGTGNGIIGLLKGETNIQSYT---KGIKVGNVTAAT 1629  
Qy 1227 --ANAGDLTVNGGAEINATEGAATLTATGNTLTTEAGSSITSTKGQVDLLAQ----- 1276

Db 1630 SPFDKAKVAIGIYADAQGTVG-----TPYNIIT-----PITAGKSGVIGIFADKDSNINYT 1679  
Qy 1277 -----NGSIAGSINAANVTLNTTGTTLTTVAGSDI--KATSGTLVINAKDAKINGDASGD 1328  
Db 1680 GNMEIGDGTTAGTGIFITKKIGATGKGVTLGTNTIKLGTGKVVAIAISEGTTFNG---GN 1736  
Qy 1329 ST-EVNAVNASGSGSVTAATSSSVNI---TGDLTNTVNGLINIISKDGRTNTRLRGKEIEVK 1384  
Db 1737 ATIELVGSNIQGVG-VYAKKSGSTVNIDHWTFFNNNGSAAEEVSESEGRVYIN-ANKNLKPK 1794  
Qy 1385 YIOPCVASVEEVIEAKRVLEKVKDLSDEERETL 1417  
Db 1795 MVLTHVINGETSIATGKTVTVSNDGSIITAKENI 1827

Search completed: March 24, 2003, 15:24:13

Job time : 86 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 15:23:05 ; Search time 26 Seconds  
(without alignments)  
5461.175 Million cell updates/sec

Title: US-10-092-880-4

Perfect score: 7407

Sequence: 1 MNKIYRLKFSKRLNALVAVS.....CFSSGNGARVCTNVADGQP 1477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	7327	98.9	1477	2 B43855	high-molecular-wei
2	4957	66.9	1536	2 A43855	high-molecular-wei
3	876	11.8	1910	2 AF0394	probable adhesin h
4	744.5	10.1	2154	2 F83068	hypothetical prote
5	636.5	8.6	1417	2 A83080	hypothetical prote
6	602	8.1	3705	2 AD0123	probable autotrans
7	593.5	8.0	4919	2 T31105	hypothetical prote
8	555.5	7.5	1577	2 A35140	hemolysin A precu
9	553.5	7.5	1018	2 H83135	probable adhesin P
10	531.5	7.2	1268	2 B99789	hemagglutinin/hemo
11	531.5	7.2	1270	2 E85649	hypothetical prote
12	523.5	7.1	4152	2 T31102	filamentous hemagg
13	520	7.0	1975	2 B81192	hemagglutinin/hemo
14	520	7.0	2535	2 AC0304	probable hemolysin
15	517.5	7.0	1995	2 G81044	hemagglutinin/hemo
16	517.5	7.0	2020	2 C48399	ABC-type transport
17	512.5	6.9	2340	2 B71704	cell surface antig
18	511	6.9	2249	2 A41477	190K surface prote
19	508	6.9	2015	2 B81989	hypothetical prote
20	504	6.8	3295	2 AE0074	probable adhesin Y
21	486.5	6.5	928	2 S54699	hemopexin-heme com
22	485	6.5	2059	2 D82671	surface protein XF
23	480	6.5	1594	2 T43072	hemolysin A - Edwa
24	477.5	6.4	1487	2 AG2560	hypothetical prote
25	475.5	6.4	1749	2 S75138	hypothetical prote
26	475.5	6.4	2021	2 F90859	190-kDa cell surfa
27	474	6.4	5291	2 F90696	hypothetical prote
28	472.5	6.4	1152	2 AE1852	hypothetical prote
29	471	6.4	2514	2 F81045	hemagglutinin/hemo

filamentous hemag  
hemopexin-heme com  
probable RTX famil  
hypothetical prote  
hemolysin (importe  
extracellular seri  
outer membrane pro  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hemagglutinin/hemo  
hypothetical prote  
ydek protein - Esc  
outer membrane pro

## ALIGNMENTS

### RESULT 1

B43855

high-molecular-weight surface-exposed protein - Haemophilus influenzae

C:Species: Haemophilus influenzae

C>Date: 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994

C:Accession: B43855

R:Barrenkamp, S.J.; Leininger, E.

Infect. Immun. 60, 1302-1313, 1992

A:Title: cloning, expression, and DNA sequence analysis of genes encoding nontypeable

detella pertussis.

A:Reference number: A43855; MUID:92192797; PMID:1548058

A:Accession: B43855

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1477 <BAR>

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBI:89237, NCBI:89240)

Query Match

Best Local Similarity 98.9%; Score 7327; DB 2; Length 1477;

Matches 1465; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLAKPLSAML	LGVT	60
Db	1	MNKIYRLKFSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLAKPLSAML	LGVT	60
Qy	61	SIPOSVLASGLOGMDVVGHTATMOVDGNKTIIRNSVDATINNKQFNIDQNMVQFLQENN	120	
Db	61	SIPOSVLASGLOGMDVVGHTATMOVDGNKTIIRNSVDATINNKQFNIDQNMVQFLQENN	120	
Qy	121	NSAVFNRTVSNQISQILGILDSNGQVFLINPGLITIGKDAIINTNGFTASTLDISNENIK	180	
Db	121	NSAVFNRTVSNQISQILGILDSNGQVFLINPGLITIGKDAIINTNGFTASTLDISNENIK	180	
Qy	181	ARNFTFQTKDALAFIVNHGLITVCGKSVNLIGKVKNEGVISVNGGSI	SLLAGQKIT	240
Db	181	ARNFTFQTKDALAFIVNHGLITVCGKSVNLIGKVKNEGVISVNGGSI	SLLAGQKIT	240
Qy	241	ISDIINPTITYSTAAPENEAVALGDI	FAKGGNINRAATIRNOGLSADSVSKSGNIV	300
Db	241	ISDIINPTITYSTAAPENEAVALGDI	FAKGGNINRAATIRNOGLSADSVSKSGNIV	300
Qy	301	LSAKEGEAEIGGVISAQNOQAKGKLMTITGDKVT	LTKTGAVIDLSGKEGETVLGGDERGE	360
Db	301	LSAKEGEAEIGGVISAQNOQAKGKLMTITGDKVT	LTKTGAVIDLSGKEGETVLGGDERGE	360
Qy	361	KGNGIOLAKTSLKSTINVSCKPKGFAIVWGDIALDGNINAGSGDI	AKTGGFVET	420
Db	361	KGNGIOLAKTSLKSTINVSCKPKGFAIVWGDIALDGNINAGSGDI	AKTGGFVET	420
Qy	421	SGHDLFTKQNAIVDAKEWLLDFDNV	SINAEPLFNNTGINDPEPTGTGASDPKNSSELK	480
Db	421	SGHDLFTKQNAIVDAKEWLLDFDNV	SINAEPLFNNTGINDPEPTGTGASDPKNSSELK	480

April 1992 after print

Db 421 SGHLSIESNAIVTKWELLDPODVTIEADPLRNNTGINDFFPTGTGEASDPKKNSELK 480  
Qy 481 TLTNTTISNVLKNAWTNITASRKLTVNSNINIGSNHLLHSHKGGVQIDGDTIS 540  
Db 481 TLTNTTISNVLKNAWTNITASRKLTVNSNINIGSNHLLHSHKGGVQIDGDTIS 540  
Qy 541 KGGNLTISGGWVDVHKHNTIDQGFNLTAASVAFEGGNKARDAANAKIAVQGTVTITG 600  
Db 541 KGGNLTISGGWVDVHKHNTIDQGFNLTAASVAFEGGNKARDAANAKIAVQGTVTITG 600  
Qy 601 EKGDFRANVSLGTGKGLNLTSSVNNLTJHNLSTINISGNITINQTRKNTSYWQTSHTD 660  
Db 601 EKGDFRANVSLGTGKGLNLTSSVNNLTJHNLSTINISGNITINQTRKNTSYWQTSHTD 660  
Qy 661 SHWVNSALNLTGANFTPIKVISSNSKGLTQYRSSAGVNFNGVNGNSFNLEKAGKVN 720  
Db 661 SHWVNSALNLTGANFTPIKVISSNSKGLTQYRSSAGVNFNGVNGNSFNLEKAGKVN 720  
Qy 721 KLPKNENMTSKPLPIREFLANITATGGSVFFDIYAHNSGRGAELKKNSEINSGANFTL 780  
Db 721 KLPKNENMTSKPLPIREFLANITATGGSVFFDIYAHNSGRGAELKKNSEINSGANFTL 780  
Qy 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNNVTLG 840  
Db 781 NSHVRGDDAFKINKDLTINATNSFSLRQTKDDFYDGYARNAINSTYINISILGNNVTLG 840  
Qy 841 QNSSSTIGNTIEKAANVTLEANNAPNOQNIQIRDRVTKLGLSLLVNGSLTGENADIKGN 900  
Db 841 QNSSSTIGNTIEKAANVTLEANNAPNOQNIQIRDRVTKLGLSLLVNGSLTGENADIKGN 900  
Qy 901 LTISESTAFKGTDRDLTINIGTNGTAEINTTQGVVKGNGVNTNOGDLNITTHAKRNOR 960  
Db 901 LTISESTAFKGTDRDLTINIGTNGTAEINTTQGVVKGNGVNTNOGDLNITTHAKRNOR 960  
Qy 961 SIIGDILINKGSLNITDSNDAIEIQIGNISOKEGNTLTSSDKINIKOITTKKIDGE 1020  
Db 961 SIIGDILINKGSLNITDSNDAIEIQIGNISOKEGNTLTSSDKINIKOITTKKIDGE 1020  
Qy 1021 DSSSDATSNANLTITKTELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEKTV 1080  
Db 1021 DSSSDATSNANLTITKTELKLTEDLSISGFNKAETAKDGRDLTIGNSDNGSGAEKTV 1080  
Qy 1081 TFNNVKDSKISADGHVNTLNSKVKTSNGSGRESNDNDPGLTITAKNVEVKNKDITSLKT 1140  
Db 1081 TFNNVKDSKISADGHVNTLNSKVKTSNGSGRESNDNDPGLTITAKNVEVKNKDITSLKT 1140  
Qy 1141 VNITASEKVTTTAGSTINATNGKASITTKGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200  
Db 1141 VNITASEKVTTTAGSTINATNGKASITTKGDISGTISGNTVSVSATVDTLTKSGSKIEA 1200  
Qy 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260  
Db 1201 KSGEANTVSATGTIGTISGNTVNTANAGDLTVNGAEINATEGAATLTATGNTLTTEA 1260  
Qy 1261 GSSITSTKGQVDDLLAONGSIAGSINAANVTLTNTTGLTVVAGSDIKATSGTLVINAKDAK 1320  
Db 1261 GSSITSTKGQVDDLLAONGSIAGSINAANVTLTNTTGLTVVAGSDIKATSGTLVINAKDAK 1320  
Qy 1321 LNCDSAGDSTEVAANVAGSGSVTAATSSSVNITGDLTVNGLNITISKDGRNTRVLRGKE 1380  
Db 1321 LNCDSAGDSTEVAANVAGSGSVTAATSSSVNITGDLTVNGLNITISKDGRNTRVLRGKE 1380  
Qy 1381 IEVKYIQPGVASVEEVEAKRVLEKVKDLSDEERETLAKLGVSARFVPPNNTITVNTQN 1440  
Db 1381 IEVKYIQPGVASVEEVEAKRVLEKVKDLSDEERETLAKLGVSARFVPPNNTITVNTQN 1440  
Qy 1441 EFTTRPSQVITIEGKACFSGNGARVCTNVADDDGP 1477  
Db 1441 EFTTRPSQVITIEGKACFSGNGARVCTNVADDDGP 1477

RESULT 2  
A43855

high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae  
C:Species: Haemophilus influenzae  
C:Date: 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
C:Accession: A43855  
R:Barenkamp, S.J.; Leininger, E.  
Infect. Immun. 60, 1302-1313, 1992  
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable  
detella pertussis.  
A:Reference number: A43855; MUID:92192797; PMID:1548058  
A:Accession: A43855  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1536 <BAP>  
A:Cross-references: GB:U08876; GB:M4616; PIDN:AAA0527.1; PID:9475771  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBI:89235, NCBI:89239)

Query Match 66.9%; Score 4957; DB 2; Length 1536;

Best Local Similarity 67.3%; Pred. No. 1.4e-194;

Matches 1046; Conservative 146; Mismatches 265; Indels 98; Gaps 20;

Qy 1 MNKIYLFKSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLALPLSAMLISLGV 60

Db 1 MNKIYLFKSKRLNALVAVSELARGCDHSTKSGSEKPARMKVRHLALPLSAMLISLGV 60

Qy 61 SIQSVLASLGDMVDVHGTATQVDGNTIIRNSVDATINNKQFNIDONMVQFLOENN 120

Db 61 SIQSVLASLGDMVDVHGTATQVDGNTIIRNSVDATINNKQFNIDONMVQFLOENN 120

Qy 121 NSAVFNRTVSNQISQLKGLDSNGQVFLINPNTIGTKDAIINTNGFTASTLIDISNENIK 180

Db 121 NSAVFNRTVSNQISQLKGLDSNGQVFLINPNTIGTKDAIINTNGFTASTLIDISNENIK 180

Qy 181 ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSGISLLAGQKIT 240

Db 181 ARNFTFEQTKDALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSGISLLAGQKIT 240

Qy 241 ISDIINPTITYSIAAPENEAVALGDIKAGGNIINRAATIRNOGLSADSVSKDSGNIV 300

Db 241 ISDIINPTITYSIAAPENEAVALGDIKAGGNIINRAATIRNOGLSADSVSKDSGNIV 300

Qy 301 LSAGEGEAFIGVISAONOAKGKMITGDKVTLTKTGAVIDLSKEGGEYVLLGDERGE 360

Db 301 LSAGEGEAFIGVISAONOAKGKMITGDKVTLTKTGAVIDLSKEGGEYVLLGDERGE 360

Qy 361 GNGIOLAKKTSLEKSTINVSKEGGEFAIVWGDIALIDGINAOGSDIAKTGGFVET 420

Db 361 GNGIOLAKKTSLEKSTINVSKEGGEFAIVWGDIALIDGINAOGSDIAKTGGFVET 420

Qy 421 SGHDLFIKDNAIVDAKWLDPDNVSIANAEDPLFNNTGINDFFPTGTG-EASDPKKNSEL 479

Db 421 SGHDLFIKDNAIVDAKWLDPDNVSIANAEDPLFNNTGINDFFPTGTG-EASDPKKNSEL 479

Qy 480 KTTLTNTTISNVLKNAWTNITASRKLTVNSNINIGSNHLLHSHKGGVQIDGDTIT 539

Db 479 KTTLTNTTISNVLKNAWTNITASRKLTVNSNINIGSNHLLHSHKGGVQIDGDTIT 537

Qy 540 ----SKGNLTISYSGWVDVHKHNTID-OGFLNITPA-ASVAFEGGNKARDAANAKIAVQ 593

Db 538 TGDDTRGANLTISYSGWVDVHKHNTIDSLGAOGNITAKQDIAFEKSGNQV-----ITGQ 590

Qy 594 GTVTITGCKDFRANVSLNGTSGKGLNLTSSVNN---LTHNLSTINISGNITINQTRK 650

Db 591 GTIT-SGNQKGFERNVSLNGTSGKGLTTRKNTKRYATNKFEGTLNISKGVNLSWLPK 649

Qy 651 NTS-YWQTSNHSNHNVSALNLETGANETTF-IKYISSNSKGLTQYRSSAGVNFVNGNM 708

Db 650 NESGYDKFKRTYNNLTSLNVSSEGEFNLTDSCRSDSAGTLTQYNNLNGISF---NKDT 706

Qy 709 SPNLKAGAKVNFKLPNNMNTSKPLP-IRFLANITATGGSVFFDIYAHNSG---RGAE 764

Db 707 TFNVERNARVNFIDKAPIGINKYSSLNYSFNGNISVSGGSDVFTLLASSNVOTPGV 766





```

Db 935 LFSPTKGLWFRGNGGTAANILLVANSYSSGE-----TVKI-NASSNKNWITA-- 983
Qy 956 KNOBSIIIGDIIINKG-SLNTDSN---NDAEIOIGNISQKGNLTSSDKNITKQI 1011
Db 984 -GKDSIIITAGSKATGPNMIENIENFTNGFTNGITWISGVNVSANGVDITSNS 1042
Qy 1012 TKKIDGEDSSSDATSANITIKTELK--LTEDLSISGFNKAEITAKGRDILT-IGNS 1068
Db 1043 GTGTGI-----VLDNTNLTITVGDINTITVNSSGKGIWIKSNSTLNSKIDITLVGS 1094
Qy 1069 NDGNSGAEBKATVTFNVKND---SKISADGNVNTLNSKVKTSSSGGREGNSDN--DTGLT 1123
Db 1095 AQONEG-----VITQSSDASRNNISAG-NITLIGKM---GNGSGOHSILNLGNVSLT 1144
Qy 1124 ITAKNVEVNDKITSKTWNTITASEKVTITAGSTINATNGKASITTKTGDISISGNTVS 1183
Db 1145 SSGRNDINGSSAGTGDVYFNV-----LNATAGNVSIAET-----KTALS 1187
Qy 1184 VSATVLDLTKSGSKIEAKSGEANTVSATGTIGTIGTISGNTVNTANAGDLVAGCAENAT 1243
Db 1188 TSLNAVLSLGGNNSIKAGNG-----WLLGKAFNTQAGAGIFRANSSLSVDNIIKGET 1242
Qy 1244 EGAATLTATG-----NLTLEAGSSIT---STKQGVDLAQAQSIAGSINAANVTLNTT 1294
Db 1243 EGVGA-FRKGIDFYCANITLIKQSLLGNGAGADTAGGAGISVTSL--AKLTVNNN 1299
Qy 1295 GTL-----TTVAGSDIKATSGTLVINAK-DAKLNQDA-SGDSTEVNAVNASGSGVTAA 1346
Db 1300 GSKMEGRSTSGTGINFPSSNNTLVFNGDGTDLIKGSSVAGTGAAGISVVNNSTGPMTIE 1359
Qy 1347 TSSS-----VNITGDLNTVNLNIIKSDGRNTVRLRGEIEVKYIQ-- 1387
Db 1360 GISDAGVHLFSAEHRIDRINTVNTGSSTHAEGRLISGNAIVDTTLTKGSINGSGVKIDS 1419
Qy 1388 -PGVASVEEIVLEAKRVLEK--VKOLSDDEERETLAKLGVSAVRFEVPNTTIT-----VN 1437
Db 1420 LPSGSV-----TRSLDNATLNGSSSGKGVETSDINGIHSSINGTTGTGNGYDIDG 1474
Qy 1438 TQNEFTTRPSOVIISSEKACFSSGNGARVCTN 1470
Db 1475 ENSNVGTGTSEADLLILOGVATTGTGIGKLN 1507

RESULT 4
F83068
hypothetical protein PA4625 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83068
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziouchi, S.D.; Warren, P.; Hickey, M.J.; Broman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2154 <STO>
A:Cross-references: GB:AE004876; GB:AE004091; NID:g9950869; PIDN:AAG08013.1; GSPDB:GN001
A:Experimental source: strain PA01
A:Gene: PA4625

```

```

Query Match 10.1%; Score 744.5; DB 2: Length 2154;
Best Local Similarity 23.9%; Pred. No. 1.1e-22;
Matches 394; Conservative 231; Mismatches 585; Indels 437; Gaps 80;

```

```

Qy 87 GNKTIIRNSVD-AIINKWOFNIDONEMVQFLOENNNNSAVFNRTVSNQISQKLGILDSNGQ 145
Db 60 KGTLLIDQSTQRAIINKWGFVSADEAVRFNPNQPGVTSSTLNRVRTACQESVIAGRISAPQ 119

```

```

Qy 146 VFILNPNGITIGKDAIINTNGFTASTLDISNENIKARNFTFEQTKDALAEIVNHHGLTV 205
Db 120 VIYNSNGVFSGSAAKVDVGLSLTTITANISDEHFQKGLIFDQPNPD-ARIVNDGSISV 178
Qy 206 GKDGSVNLIGKVKNEGIVSVNGSSISLLAGQKTTI---SDIINPTIITYSAAPEA- 260
Db 179 AEKGLAAFAVAPSVAANNVGNARLGTVMAAGNAATIDLDGDLVSTAVTPVTRKPDQAQ 238
Qy 261 ---VNLGDIFAKGNINVRA---ATIRNOGLKLSADSVSKO---KSGNIVLSAKGEAEI 310
Db 239 ALVNSGALQADGQSVLITAEQASRVVDNANVLSGVLARGTEVRGSSVALVSKSDIQI 298
Qy 311 GGVISAQNOQAKGKLMITGDKVTKTGAVIDLSG-KEGETYLLGDEGEKNGIQLAK 369
Db 299 AGKIDVSGPK-NGCDVLVSGQVALASTASIDARGTAQGSVRIGDFQGRGE--LPRAK 355
Qy 370 KTSLEKGSITNVS--CKEKGGAFAVWGD--IALIDGINAOGSGDIAKTGGFVETSGH-DL 425
Db 356 NATLAKGASIDVSATCKGNGGLAVVWSDGNTRMQGRILARG-GAOGNGGLVETSGKVL 414
Qy 426 FIKDNAIV-----DAKEWLLDFDNVSIADPFLFNNTGIDNDEFTGTGEASDPKKNSE 478
Db 415 SIADSAVSVAAPIYNGGTWLLDPTTLRIVA-----SGTSGSVGGANGASGDATVNAS 468
Qy 479 LKT-TLTNTTISNYLKNAMTMITASRKLTVNSSI---NIGSNH-LILHSGQREGGVQ 533
Db 469 VVTGALAGGKVT-----LSASDRLSVEAPLITSNLGGASRGLELIATGP-AGAVD 517
Qy 534 IDGDIITSK-----GGNLTIVSGGVWVDVHKNTILDQGLF---NITASVAFEGGNKA 582
Db 518 ISAPILFRNGSLAIRAGGNINFLSGGTPQTSIGIVDLGSGTLMQSTAGKISQOAGT-AA 575
Qy 583 RDAAN-----AKIVAQCTVITTEGCKDFRANNVSLNG-----T 615
Db 576 LIAANLAGRAGSIDLASWDNYAGNALQFTNGTLKVRQSNATGVITSGTVDFPFTNQST 635
Qy 616 GKGLNIISSVNN--LTHNLSGFINISGNTITINQTRKNITSY-----W 655
Db 636 GTAQNIYVSGVTRILEANSVGI---TGNITL--TAGNSEFDRLVFTALPYRRVSGSASF 690
Qy 656 QTSHDHNVNVSALNLE-TGANFTFK-----YISSNSKGLTT---OYRSSAGV-NFN 702
Db 691 PTNDSYDLVTLNRYOVNGSNVTATPNGAPSGFTVAAGNGSVTTWTGNGWTSWVKGF 750
Qy 703 GYNG-----NMSFNLKEGAKVNFKLPKNENN----- 728
Db 751 GVGIVTDELQYDVGTLTBEILIFGLGKTSRVDTRLDLDFMRGAFNFAERAQVEMFKTT 810
Qy 729 -----NTSKPLPIREFAN-----ITATGGSVFFDIYANHSRG---AELKMS 768
Db 811 TTAGDILSRQQTATILTANDATRVYGDVNPFLTATWSGINADAYVNSQFNDLYQATAST 870
Qy 769 EINISGANFTLNSHVRGDDAFK-----INKDLTINATNSFSLRQTKDDFYDGYARNAI 823
Db 871 ATQASNVQYAITNANGSEYFSQRYQLVRQDRLTVPQALIVSADAKTKVYGDADPTL 930
Qy 824 NSTYNIS-----ILGQNV-TLGGQSSSSITGNITTEK-----AANVTI----- 861
Db 931 --TYQSLKNSDSTAAGVLSGNLGRVAGEN-----VGNVGIQGLGLGNTANTYLLSYGN 983
Qy 862 EANNAPNOQN-TRDRVIKL-----GSLLVNGSLN-LTGENADIK 898
Db 984 DLURITPAQLNVJADAKTKVYGLDLPALTYQVYSGLKRGTAGAVLNGGSLSRVAGENVCV 1043
Qy 899 GNLTISESATFKGTRDTINITNGFTNNGTAINITQGVVKL-----GNVTNDGDLNITT 953
Db 1044 G---INQGLGLVSSNYTLNQQN-----NLITKALLNVJADAKTKVYGDADPALTY 1093
Qy 954 HAKRNORSIIIGDIIN-----KKSNIITDSNNDAEIOIGGNISQK 996
Db 1094 QVSGLKNGDTAGAVLNGGSLSRVAGENVCVYGINOGGLLSANVYDLSYQ--GN----- 1145
Qy 997 NLTISSDKINITKQITIKKIDGEDSSSDAISNANLITKTELK-----LT----- 1042

```

Db 1146 NLITKALLNVIADAKTKVGDADPS-----LTYQVSLKNGDGTAGSILUTGLNRA 1196  
 Qy 1043 --EDLSISGFKAEITAKQGR-DLTIGNSDG-----NSGAPAKTIVTENNVDKSKI--- 1090  
 Db 1197 AGENGVYGINQDGLALNSGNYDLVSGNNLITKALLNVIADAKTKVYGDADPSLTYOV 1256  
 Qy 1091 ----SADGHNVTLN--SKVKTSSNGGREGNSDNDTGLITAKNVEVNDITSLKTVNIT 1144  
 Db 1257 SGLKNGDTAGVNLNGGLVRVSGENVGYAIOQGGUGL-----VSGNTDL-AYOENNLIT 1309  
 Qy 1145 ASEKVTYTTAGSTINATNGKA--SIT-----TKTGDISTGTI-----SGNTVSV----- 1184  
 Db 1310 ITRKALLNVIADAKTKVGDADPSLTYQVSLKNGDSAGSILTGGLNRAAGENVGYGINQ 1369  
 Qy 1185 -----SATVDLT-----TKSGSKI--EAKS-----GEANVTSAITGIGTISGNTVNV 1225  
 Db 1370 GDLALNSGNYDLVSGNNLITKALLNVIADAKTKVGDAD-PSLTYQVSLKNGDGTAGA 1428  
 Qy 1226 TANAGDLTVNGGAEIN--APEGAAITLTATGNTLTTEAGSSITSTKQGVDLA----- 1275  
 Db 1429 VLNGGLVRVSGENVGYAIOQGGUGLVLGSGNYDLAYOENNLITKALLNVIADAKTKVY 1488  
 Qy 1276 -----QNGSIAGSINAANVTNTTGLTTLTVAGSDIKATSGTLVINAKDAKLN 1322  
 Db 1489 DADPSLTYQVSLKNGDGTAGV-----LN-GGSLSKVAGENV-----GVYGINOGDLALN 1537  
 Qy 1323 GDASGDSSTEVNANVNASGSGVTAATSSSVNITGLNLT-----VNLN----- 1364  
 Db 1538 ---SGN-----YDLSYOGNNLITKALLNVIADAKTKVYGDADPSLTYQVSLKNGDRA 1588  
 Qy 1365 --IISKDGRNVTLRGKEIEVKYIQPG 1389  
 Db 1589 GAVL--NGGGLVRVSGENVGYAIOQG 1613

RESULT 5  
 A:3080  
 hypothetical protein PA4541 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text\_change 31-Dec-2000  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 -; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: A83080  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1417 <STO>  
 A:Cross-references: GB:AE004867; GB:AE004091; NID:9950769; PIDN:AG07929.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4541

Query Match 8.6%; Score 636.5; DB 2; Length 1417;  
 Best Local Similarity 21.9%; Pred. No. 1.6e-18;  
 Matches 352; Conservative 238; Mismatches 523; Indels 495; Gaps 72;  
 Qy 1 MNKYLKFKRLNALVAVSELARGCDHSTKGESEKPAKMKVRHLALPLSAMLISLGV 60  
 Db 1 MNKSYTLWNQ-----ATGCWNVASEGTRRSK-SGRCKALVVGASILLGL-FC 47  
 Qy 61 SIPOSVLASGLOGMDVVGHTATMQ--VDGNKTLII-RNSVDALINWKFNDONEMVQFLO 117  
 Db 48 QAPAFALPS---GATVVGSDGFGFTSDGRHVVLDQGSKHLITWNFEFSVRADERVSHQ 104  
 Qy 118 ENNSAVFNVRVTSNQISOLKGLDSGOVFLINPNTIGKDKATINTNGFTASTLDSNE 177  
 Db 105 PQDAVALNRVIGRNSDIOGRIDANGKVLVNPNGVYFGKSAQVNVGGLVASTLIDADR 164

Qy 178 NIKARNTFEOTKALAEIVNHGLITVKGDSVNLIGGKVKNEGVISVNGSISLLAQ 237  
 Db 165 DFLAGNYQFSGDSG---ATVSNAGSLQASEGSGIALLGARVSDNGLLQAOGLDVALGAG 221  
 Qy 238 KITSDIINPTITYSIAAPENAVNLGDFAKGGINVRA-----ATIRNOGKLSADSV 291  
 Db 222 GINLN-----FDGDGLNLNQVDKGSVDALAHNGGLIRAD-- 255  
 Qy 292 SKDKSGNIVLSAKBEAEIGVISNON-----OAGGKGLMITGDKVTILTKTGAVIDLS 344  
 Db 256 ---GGQVLMARSADSLTKTVVNNQGLTEARTLSAEGRIVLDDGEGOTVRAVGAKQDAS 311  
 Qy 345 KGEGETYLGDEREGKNGIOLAK---KYSLEKSGTINVSQKKGKGFALVWGDIALID 400  
 Db 312 AIGGNG---GGLVLNQGAG-VEIORTAQVDTHADGAT-----GTWRILSHSVSVA 359  
 Qy 401 -GNINAOSGDIATGGFVETSGHDLFIKDNAIVDAKEMWLLDFDNVSIINAEDPLFNNGTI 459  
 Db 360 VQANAAAGD-----SGOVHVAQGPAGANAS---DSNGVTIVOOOP----- 397  
 Qy 460 NDEFPTGTGEASDPKKNSLKTTLTNTTISNYLKNATMTNITASRKLTVNSISINISNSH 519  
 Db 398 -----AVDLAAGANGTSAVOSQSGANGISVVQSO-----NSPNIGSGAN 442  
 Qy 520 LTLHSGGORGGV-----QIDGDTISKGNLTIVYSGGWVDVHKNTILDQGFNL 567  
 Db 443 GLSVVOSQGANIGAGSGISVVOSONSPNIGSGVNGVTV-----VQSONGANIGSGAG 497  
 Qy 568 ITAASVAFEGGNKAKDAANAKIYAGQTVITTEGKDFRANNVSLNGTKGLNIISSVNN 627  
 Db 498 IIV--VOSONGANIGSGAGISVVOSQSGPSIGSG---VNGVTIVQSONGANIGPGVSG 551  
 Qy 628 L-----THNLSGTINISGNITINOTTRKNTSYWQTSHD---SHWNVSALMLETGANPT 677  
 Db 552 IDVQVQTLPLNLSPGANGSSIVQV-----QTLPDIAADAGNVHVVOQTGGNKV 600  
 Qy 678 FIKYISSNSKGLTQYRSSAGVNFNGVNGMNSFNLKEGAKVNFKLKPNENMTKPLPIR 737  
 Db 601 F-----GNSATNYSR---TVQARSNENVSG----- 624  
 Qy 738 FLANITATGGGSVFFDIYANHSGRAELKMSINISNGANFTLNHSH----- 784  
 Db 625 -LANPSSAGKGST---LHADTLARLSTSNVEVATRG-----NAHVAPLSDWSGNGLT 675  
 Qy 785 ---RGDDAFKINKDLTINATSNFSLROT KDDFYDGYARNAINSTYINISILGGNWTLLGG 840  
 Db 676 LTAERGD--LRINGALTQAGENASLTNA-----GORPLRIDDSLSLTGOCARVEF-- 724  
 Qy 841 QNSSSITGNITIEKAANVTLEANNAPNOCTRDRI-----KLSGLLVNLSL 888  
 Db 725 -NSDK---GYALAEGRITLSGKNAGFRANGRYSVIOQLQDLRGIDRLDLSGVN- 778  
 Qy 889 SLTGENADIKGNLTISESATF-----KCKTDTDLNI--TGNET-----NNGT----- 928  
 Db 779 RIAGGNSSF---LSYGNASAFGCTFDGLNTIDNLAVYGTGAYSGLFVNRGTLRLNLE 835  
 Qy 929 -----AEINITOG---VVKLGNV--TNQDGLNITHAKRNORSIIGDIIKNGSL 974  
 Db 836 RISADGAQATHYVQVGSIAAVNLGRIDNVNASDIRIAAASKLNS---LGGVLVALNLSGI 892  
 Qy 975 NITDSN-----NDAETIQGG---NISQKEGNTLTSSDKINITKQITTKKGDGEDSSD 1025  
 Db 893 DNASAGTLVGNRHTYALGGLAAENISTARGVASISNSR----- 931  
 Qy 1026 ATSNANLTIKRELKLTEDLSISGFKAEITAKDGRDLTIGNSNDQ---NSGAEAKTVTF 1082  
 Db 932 -----ADFAISGQLK-DHASYGAGGLVGRNGLIRSSQSO----- 967  
 Qy 1083 NNVKDKSIADGHNVTLNSKVKTSSNGGREGNSDNDT-----GLTITAKNVEVNDIT 1136  
 Db 968 ---CGLSLUSGHMGLGVYSGAGGLADVASVVDVSGNCGORGLYGLIGLGNVSGI- 1021  
 Qy 1137 SUKTVNITASEKVTTT-----AGSTINATNGKASITTKTGDIS-----GTI--- 1177

```

Db 1022 ----AHATAGKVRGTDAEALGGLGRLNNAALNNAAS----AHGDSVLAQGRVGLGLGNH 1074
QY 1178 -SGNTVSVSATVLDLTKSGSKTEA-----KSGEANTV-----SATGTIGT 1217
Db 1075 OAGNLANVSTGNLS--GGSLQAGGLGLNANASVLNASAKGNVATRGAEAVGGLGN 1132
QY 1218 ISGNTVNTANAGDLTVNGCAINATEGA-----ATLTAT----- 1252
Db 1133 LVGSVINGSA-SGEVDSGKTLGGLGNSLGNHNSLKASGVWAGANSVDVGLGHNH 1191
QY 1253 -GNVLTTEAGSITSTKGVQVLLAONGTAGSINAANTV-LNNTCTLT-----VAG 1302
Db 1192 GGNHSTLAASGNVTKGK-----SRVGLGVYNDAASTNVASGNVSASGRATGGLIG 1246
QY 1303 SDIKATSGPLVINAOKDAKLNCDAS-----GDSTEVAVNASGS----- 1340
Db 1247 SCLR--GSLMLASSHGIVNDKTSINLGLVGRGENTSIKAKASCAVSGGAGIRAGGLV 1303
QY 1341 -----GSVTAATSSVN--ITGDLNLTVNGLNILSKDG 1370
Db 1304 GSLEGWQALILGASAGDVTAGYDSYIGGLVGFSTATISGASASCKVG 1351

RESULT 6
AD0123
probable autotransporter protein yaph [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0123
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
J. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0123
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3705 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89847.1; PID:g15979073; GSPDB:GN00175
C:Genetics:
A:Gene: yaph

Query Match 8.1%; Score 602; DB 2: Length 3705;
Best Local Similarity 21.7%; Pred. No. 1.4e-16;
Matches 413; Conservative 260; Mismatches 695; Indels 534; Gaps 91;

QY 14 NALVAVSELA--RCGDHSTKGS--EKPARMKYRHLAKPLSAMLISLGVTSIPQSVLA 68
Db 972 SGVTVTDALNSGVNTIGSSGNDNDPLATNVSLLEQDRAGSTLELINATNVTDIDA 1031
QY 69 SGLOGMDVVHCTA-TMQVDGKNTIIRSVDAIINWKQFNIDQNMVQFLOENNNNSAVFNR 127
Db 1032 LDL-----LVNGTALTSGTGVSQSAIQGGSTVA-----NATHYGLASSNSNGDGLYVN 1082
QY 128 VTSNQISQLKILSDNSGVFLNPINGITIGKDAIIN-----T 164
Db 1083 YT---LSALELLADGADALLATATESGLTANR--VLNNAELFVGLGVVDAQNGALTLANGS 1137
QY 165 NGFTASTLDISNENIKARNETFEOTK-----DKALAEIVNHGILITVKGOG 209
Db 1138 NRYEGTTTAVTAGELILGANGAFGCTSLDIDASGANINCYSTVGAVTVNGVTVLGSGL 1197
QY 210 SVNLIGKVKRNEGVISVNGGSIISLAGOKITISDIINPTITYSIAAPEAVNLGDI FAK 269
Db 1198 V--LTSGLTNGGLTDLTGALNLTAGASTVAGGLTGAGTLNI-----N 1240
QY 270 GGNINVRRAATIRNOGLKS-AD--SVSKDKSGNIVLSAKE--GEAEITGGVISAQNOQAKG 324
Db 1241 GGNLSVSAANSGLSGOIHADVASVLTIDTGLTSAVEVLGTLNLNGANAATN----- 1295
QY 325 KLMTGDKVTLKTCVADILDSKE--GGEYTLGODER----GEGK----- 362

```

```

Db 1296 ---VLSGDG-TINTNAAVTLSSGNSFSGAHQITGDTGELTVGOASNLGASSATVNLGLTSH 1352
QY 363 ---NGI--OLAKTSLKSGSTINVS-----KEGGFAIVMGDIALI-----DG 401
Db 1353 LILNGVSESIANVLISGVAGSTVDIIGGADTALTANNSGFL---COYALAGNSKLTVASTN 1409
QY 402 NINAQ-----GSGDIAGTGGFVETSGHD-----LFKIDNAIVD-----AKEWLLDF 442
Db 1410 NLGASSVALAGAGDTLSLGSFNCTFGNSVTSVGLQVTDADAEVLTSSNGVSNVITDI 1469
QY 443 DNVSINAEI-PLPN-----NTGIND-----EPTGTGEASDPKKNSELKTLTNTT 487
Db 1470 ADATLNLDDIALEFNHVLTNGLLNVAKNDASTAFDFGTVGGAFSGIVN-----LTNTT 1523
QY 488 ISNVLKNAWMTNITASPKLVNSSINIGSNHLLILSKGORGQGVQIDG----- 536
Db 1524 FALSADNAAL-ARATLKLSDSVTVTVGATDR-TLHGLDLNGTGLIFDGSPPSQANGVV 1581
QY 537 ---DITSKGNLTIY-SGGWVDVH-----KNITL---DQGFNLITAAVAFEGGNKARDA 585
Db 1582 TVTDLALNSGTISITGAGNWNENHPVTPPNVSLLEODRGDILLELINAANVTGN-----A 1636
QY 586 ANAKIVAQGTVTITG-----EGKDFRANV-----SLNGTG----- 616
Db 1637 NNLDLVDGTAITSGTQGVESAIOQGGSTVANAHHNYGLTSSNGGSGLYVNTLSALE 1696
QY 617 ---KGLN--IISVNNLTHNL---SGTINISGNITINQTRKNKTSYMWOTSHDSHNVSAL 668
Db 1697 LLANGANALLLATESGLTANRVLAELFGVGLVDAQAQALTLANGNNRYEGTTVTAG 1756
QY 669 NLETGANFTF---IKYISSNSKGLTQYRSSAGVFNPG-----VNGNM 708
Db 1757 ELILGANGAFQTSLLNIAASGANINGYRQTVGAVTNSGAVTLGNGVLTSCLLTNGGI 1816
QY 709 -----SFNLKECAKYNF-----KLKPN-----ENMNTSKPLRPLANTATG 746
Db 1817 LDLTGGLNALAAGSSSTVAGGLTGAGTLNGLDGLAVSATNSGLSQTHIADVASVTLTG 1876
QY 747 GGSVFEDIVANHSGRCAELKMEISINSGANFTLNHVRGDDAFKINKOLITNATNSNFS 806
Db 1877 TGTL-----GTSAVEVLGTLNL--NGANAAMTVLSSGGVINTNAAVTLSGNSNFSG 1926
QY 807 LRQ--TKDPFDGYARN--AINSTYINISILGNNVTLGONSS-----SSITGN-ITIEKA 856
Db 1927 AHQIGTGDGELTVGOASNLGASSATVNLGLTSHLILNGVSESIANVLSGAVAGSTVDIIG 1986
QY 857 ANVTLEANNAPNOQNIIRDRVILKISLLVNGSLSLTGENADIKGNLTISESATFKGKTROT 916
Db 1987 ADTALTANNSG-----FLGOYALAGNSKLTVASTN---NLGASSSVALAG-AGDT 2032
QY 917 LNTGNFTNNGTAEINIT--QGVVVKLGNVTNDGDLNITT-----HAKRNORSI 962
Db 2033 LSLSG---ENGTFGNSVTSGVLQ---VTDDAEVTLTSSNGVNTVKVDIADATLNLNDI 2086
QY 963 IGGD-IINKKGSINITDSNNDAEIQIG-----GNISQKEGNETISSDKINITKQITI 1013
Db 2087 ALFDHVLTCNGTLNVAKNLATTAFDFGTVGAFSGIVNLTWTFALSADNAALARATL 2146
QY 1014 KKGID-----GED-----SSDATSNANLTKTKELKITE-DLSISGFN 1051
Db 2147 KLSDDSVTVTGTTDRILHGLDLNGGTLIFDGSPPSQANGVVTVTDLALNSGTISITAG 2206
QY 1052 KAE-----ITAKDGRDLTI-----GNSNDGNSGABAKTVTFNNVKDSKIS 1091
Db 2207 WNEHEPVTTPPNVSLLEODRGDILLOLIDADNVNTGNANDLELMINGTITISAGQGVSTVQ 2266
QY 1092 ADGHNVTLSNKVKTSSNGGR-----ESNSD-----NDTGLTITAKNVEVN 1132
Db 2267 OGYTVANATHNYGMTSNGGSLVYNYTLTSALELLADGANALLLATESGLT---ANRELN 2323
QY 1133 KDITSILKTVNITASE-----KVTTTAGSTINATNGKASITTKTGDISGTIS 1178

```

Db 2324 AELSGVGLVDAQNCALTLANGNNRYEGTIVTAGELILGANGAFGOTSLNIAASGA-S 2382  
 QY 1179 GNTVSVATVDLTTSKSGKIEAKSGE-----ANVT-----SATGTIGGTI 1218  
 Db 2383 ANINGRYQTAVGTGVTGLNGELISTDFLINTGMINVTDGILNLENGGASSISGL 2442  
 QY 1219 SGNVTNVTANAGDLV-----CNGAEINATECAATLTATGNTL--TTEAGSITSTKQV 1271  
 Db 2443 TNGI-LNTKGGDTISIDNGLAGTNIISDCASTVLNGGTLTGCTIGTNGLSGVIVDLGL 2501  
 QY 1272 DLLAONGSTA-----GSIN-AANTVLA-----TTGTLTAVAGSDIKATSGT- 1311  
 Db 2502 NLVADN-SLANVSGDTTINTATVTLSGNSSFSFGAHQIGTNGELTVQASMLGASSATV 2560  
 QY 1312 -----LVINAKDAK-----LNGDA-----SGDSTEVAVNA-----SGSGSVT 1344  
 Db 2561 NLGTLTSHLILNGVSESIANVLSGVAGSTVDIIGGADTALTANNSGFLGOVALAGNSKLT 2620  
 QY 1345 AAT-----SSSVNI--TGDLNTVNGLN-----IISKDG-RNT 1373  
 Db 2621 VASTNNLGASSVALAGTGDTLSLGFGNCTFGNSVTGCVLOVTDAAEVLTSNGVNA 2680  
 QY 1374 VRLRKEIEVYIQPGVASVEEVIEAKRYLEKVKOLSDDEERETLAKLGSVAVRFEYPNNT 1433  
 Db 2681 VTIDIADATLNL--DDIALFNIALTGNGLLNVAKNDASTAFDFGATVGGAFTGTYNLANS 2738  
 QY 1434 ITVNTQNETTRPSSOVIITSEKACFSSGNGARVCTNVADGG 1475  
 Db 2739 TFDLSGNNTVLAQATLKLSSGNLT-SVNGVQVNTGLAMNG 2779

RESULT 7  
 T31105  
 hypothetical protein 2 - Haemophilus ducreyi  
 C:Species: Haemophilus ducreyi  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
 C:Accession: T31105  
 R:Ward C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.  
 J Bacteriol 180, 6013-6022, 1998  
 A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.  
 A:Reference numbers: 220984; PMID:99030326; PMID:9811662  
 A:Accession: T31105  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4919 <MAR>  
 A:Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1  
 C:Genetics: IspA2

Query Match 8.0%; Score 593.5; DB 2: Length 4919;  
 Best Local Similarity 23.0%; Pred. No. 4.4e-16;  
 Matches 385; Conservative 226; Mismatches 604; Indels 459; Gaps 83;

QY 2 NKTYLRKFKRLNALVAVSELARGCDHST-----EKSGEKPARKMVR-----H 44  
 Db 3 NKRYKLIFSFKVKNCLVPVAENIKASGNSGSSSKIAEDQEEPDLSACLSPLSSSIH 62  
 QY 45 LAL-----KPLSALLSLGVTSIP-QSVLASGLQ-MDVYVHGATMQVD----- 86  
 Db 63 LGLHNSPLKVFKGKLSVYLLSL---MPAMQVWADSSNAIVDHSHGAKQTAVDERDPK 118  
 QY 87 -GNKTIIRNSVDALIN-----WKQFNI-----DONEMVQFLOENN 120  
 Db 119 NGREKV-----VINIKPDEQISDNHFSKENIPNSAVFNNS IREKNSQLVGLLGENK 172  
 QY 121 N-----SAVENRVTNSOISQLGILDSNGO---VFLINPGLITCKDAIINTNGFAST 171  
 Db 173 NLGSOAKTIFNOVTDGQESKISGLLEVGEKADLFIINPGLVINCVTINTDREAVST 232  
 QY 172 LDISNENIKARNFTFEOTKDALAEIVNHGLITVKGDK-----SVNLIGGKVKNEGVI 224  
 Db 233 SEVVEPHIKQLN-----VQRGKVIIGDKGVATNGLSHFIDVVAKKNIEQOGKV 278

QY 225 SVNGGS-----ISLLAGOKITISDIINPTITYSTAAPEAVNLGDIKFAKGNINVRA 277  
 Db 279 SIEGDSKPAKLANVTEAAG-----NLTYD-----VNRDVT-----NRNT-- 312  
 QY 278 ATIRNOCKLSADSYSKDSKGNIVLSAKGEAGIEGGVISAQNOOAKGKMLMTIGDKVTLTK 337  
 Db 313 ---NPKPIPTDNPRK--NIAIS---GE-----SAGSMYGRNIKFIVT----- 347  
 QY 338 GAVIDLKSGEGGYLGGDERGCKN--GLQAK-----KTSLEK--GSTINVS 383  
 Db 348 -----DKGAGVNHQGVIFAEDDINILTDGNSRLNKVADYVRVVG 388  
 QY 384 KEGGPAIVGWDIALI--DGNINAQSGDIAKTGGFVETSGHDLFIKONATVDAKELLDF 442  
 Db 389 K-----DIELANNGQIHADQOQLLNATGHVKLNDGSSVISNNLIGSALN--LTL 436  
 QY 443 DNVSINAEPLP---NNTGINDPEPTGTGEASDPKNSL-----KTLT---TWTTI 488  
 Db 437 ENATVSANLFRVTDTKLNNLSKVARSADLOSGLNLDKASVLAHLKLTILNISDVSL 496  
 QY 489 SNYLK-NAWTNITASRKLTVNSS-----INIGSNHLILHSKQKRGG-----V 532  
 Db 497 NNQKLSANNLIKIKKVRDLNINNSLSANNLTLSNNTILKNKSKFTAGMTLNVNNTV 556  
 QY 533 QIDGDIKSGGNLTYSGGVVDVHKNTLDQGF-----LNIT-----AAS 572  
 Db 557 TLNNDSELAANLTL-----NVTKNVTLNDASKLSANKLDLNTVNTLNSKSTLSAGE 610  
 QY 573 VAFEG-----GNKARDAANAKTIVAOQTVITGEK-----DFRANNVSLNG-----TG 616  
 Db 611 LTFKKVKNVTLNNDSELAANLNSHNVTLNNKSKLSAQKADIKAVNLTLNDTTELTA 670  
 QY 617 KGLNIISVNNLTNLSGTI--NISGNTINQITRK-----NTSYWOTSHDHSW-NV 665  
 Db 671 KNLDI-----NSTTITNCTIAGIFANITTEKLNKKEKALILAEONLNF--TVNGSHVENK 724  
 QY 666 SALNLETGANETFK--YISSNSKGLTYQVRSAGVNFNGVN-----GNMSFNLKE 714  
 Db 725 GDIVSKDKATVFSKNSDFTSNGSKLVNAQNLKVVNNVNETISQDDITLIGNVTLN-AS 783  
 QY 715 GAKVNFKLKPNENMTSKPLRFLANITATGGGVDFDIYANHSGR---GAEKKSEI 770  
 Db 784 GTFTN-----SGNLTVTKLVGDQIONFTKNGLTVGEDLHKSKTKITNDGKLISKNL 838  
 QY 771 NISGANFTLNSHYRCDAPKI-----NKDLTINATNS--NFSLOTQKDDFYDGYARN 821  
 Db 839 NISSADFINNGTLGIEALKIATKGNFTNKEKAILASNSLLDTSVAGSKTFNG----- 894  
 QY 822 AINSTYINISILGGVNTLGGQNSSSITGNITIEKAAVNTLEANNAPNOQINRDRVILGS 881  
 Db 895 TIESGKNLNTNWTGAFLVNDNATIRSGVLNITSTGV---SNN-----GT 937  
 QY 882 LLVNGSLSTGNADIKNLITISESATPKGKTRDTPLNITGNFTN--NGTAEIN-ITQGVV 938  
 Db 938 LISNERLNTS-----AANFTNESGTVSNGLNIIIA 970  
 QY 939 KLGNVTNDG-----DLNITTHAKR--NORSIIGG-DIINKKGSNITDSN---NDAETQ 986  
 Db 971 KOGNTNKNLTIASROOLNLTAVADNITNDSNISKNIAVLHLSGLNLSLAKSQOYVNLGEY 1030  
 QY 987 IGGNISOKEGNLITISSDKINIKQITIKK-----GIDGEDSSSDATSNNAN 1031  
 Db 1031 AGNNISVRAHQL---KNDVKLMGDITTKTKEGQASYKLYQASNGHFGNDGSS-GYSEGD 1086  
 QY 1032 LTIKTKELKLTEDLSTSGFNKAEITAKGDRDLTIGNSNDGNSGAEAKTVTFNNVSKDIS 1091  
 Db 1087 LNIKGFADLKNLTQVRIK--IYA--GRDLTFNKSNAAGK-----SEII 1128  
 QY 1092 ADGHNVTNLSKYKTSSSNGGRESNDNDTGLTITAKNVEVNNKDITSLKTVNITASEKYVT 1151  
 Db 1129 NRG--TINVKNKLSVDSVFPENNMQSOKVDLYTKIFEAKSDIE--LTFKNGTHPYVL 1183  
 QY 1152 TAGSTINATNGRASITTKTKGDISGHTISGNTVSVSATVDLTT-KSGSKIEAKSGEANVTSA 1210

```

Db 1184 NFKSNNEKRYRSENTEKFNKSGIDGDLNEALSUSAPAEIAEYSGS-----SSNVINPVSY 1239
QY 1211 TGTIGTGGTGNVNTANAGDLTVGNG--ABINATEGAATLTATGNTLTTEAGSSITSTK 1268
Db 1240 LAALGNANNSNPHVNTALFKHILGNGWQDDLKQENIKVLKQKWEDEKDKDGAS----- 1294
QY 1269 GOVDLQAQNGSIAGSINAANTVLTGTLTVAGSDIKATSGTLVINAKDAKLN-GDASG 1327
Db 1295 KWLIDLPNTDEKAKI-PAGIIRNGNDTISDVESEDEKFKYSKFQ-NGEAKKNDTGDSY 1352
QY 1328 DSTEYN-----AVNAGSGSVTAATSSVNITGDLNTVNGLNIIISKDGRVTLRKG--E 1380
Db 1353 DSTKASEKRYKVENVDKENI--DEHKLNIGKHEITVPGVSPENLNKKNMDHOPDKLGE 1409
QY 1381 IEVKYIOPGVASVEVEIEAKVLEKVD--LSDEERETLAKLGVSAVRFVEPN 1432
Db 1410 IDKSIISELLA--OPVYTEKSAARDSPRVQNQDKREALONLYKTRLSYINQNN 1460

RESULT 8
A35140
hemolysin A precursor - proteus mirabilis
C:Species: proteus mirabilis
C:Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 08-Oct-1999
C:Accession: A35140
R:Upjohn, T.S.; Welch, R.A.
J. Bacteriol. 172, 1206-1216, 1990
A:Title: Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin gene
A:Reference number: A35140; MUID:90170827; PMID:2407716
A:Accession: A35140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1577 <UP>
A:Cross-references: GB:M30186; NID:gl50888; PIDN:AAA25657.1; PID:gl50890

Query Match 7.5%; Score 555.5; DB 2; Length 1577;
Best Local Similarity 20.7%; Pred. No. 3.7e-15;
Matches 361; Conservative 257; Mismatches 606; Indels 523; Gaps 79;

QY 40 MKVRHLAKPLSLGVTSTIPOSVLASGL-----QGMQDVV---HGTATMQV---DGN 88
Db 1 MKNFNLSPSGRLAASLAIFVLSNAYNGIYVDPDAGHQPVDVSAVNGQVQVINTVTPNN 60
QY 89 KTIIRNSVDIAINWKQNI-----DONEMVOFLOENN-----SAVENRYT 129
Db 61 EGISHN-----QYQDFNVGPKGAVFNNALEAGQSQLAGHLNANSNLNGQAASLIINEV 114
QY 130 SNOISQLKG---ILDSNGOVFLINPNCITIGDKDAINTNGFTASTLIDISNEIKARNTF 186
Db 115 SRNPSFLGQOEVEFGAAEYVLSNPNGITCDGCGFINT-----SRSSLVVGNPULF 164
QY 187 EOTKDALAEIVNHLITVGKD-----GSVNLGGKVKMGVSVISNGSGISLLAGOKITI 241
Db 165 ENGOLKGYSTLNNTNLISLGNLNTGLDLPAPRISRG--KITAAEISAFITGON-TF 221
QY 242 S---DIINPTIYS-----IAAPENAEVNLGDIAPKGNINVRATIR 281
Db 222 SQHFDILSSOKPVYALDSYFFGWSQGRIRINTAGSGVKLAGKTAANDLSVRADNIQ 281
QY 282 NQKLSADSVSKDKSGNI-----VLSAKEGEAEIGGVISAQNOQAKGK 325
Db 282 TDSQRYDSYDKDGSQNYQYRGITVNNSGSGSQTTLTKELKGNITLVASSHNQIKASD 341
QY 326 LMITGOKVTLTKTGAVIDLSGKEGGEYLVGGDERG--EGKNGIQAKKTSLEKGSITVNSG 383
Db 342 LM--GGDIITLQ--GADLTIDGKLOQKQETIDNRWFYSWYDYTKKEQIQIGSQID--A 396
QY 384 KEKGGFAIVWGDIALIDGNA-----QSGGDIAKTG----- 415
Db 397 KKNATLTATKGDTLDAKINAGNNLAINANKDIIHNLGLVEKRSSENGKRNHTSRLES 456
QY 416 -----GFVETSGHDLFIKIDNAIVDAKEWLLDFDNYSI 447

```

```

Db 457 GSWNSHOTETTLKASBELTAGKDLGLDAQGSITAQGAKLHANENVLYNAK-----DNINL 510
QY 448 NAEPLFNNTGINDFPTGTGEASDPKKNSEKLTTLTNTTISNYLKNAMTMTASRKL 507
Db 511 NVQ-KTNNDKTVTDNVHVMGGTGGGKNNNNQOQVSHAT-----QLTADGQLL 558
QY 508 VNSIINGNSHILHLSKQORG-----GGVQIDGDIITSKGNLTIIYSGGVWDV---H 556
Db 559 LAADNNVNITGSQV---KNGQAFVKTQGVVDNALSETISKIDERTGTAFNITKSSH 615
QY 557 KNITLDQGLFNITAAVAEFGGNKARDAANAKIYAQGTVTITGEBCKFRANNVSLNGTG 616
Db 616 KNET-----NK-OTSGSELISDAOLTUVVS-----GNDVNVIG-- 647
QY 617 KGLNISSVNNIETHNLSGTINI-----SGNITINQ 646
Db 648 ---SLIKSADKLGIIHSLGDIIVKSAQOVTKIDDEKTSLAITGHAKEVEDKQYSAGFHITH 704
QY 647 TTRKNTSYWQTSHDSHNVVSALNLETGAN--ETF-----IKYISSNS 686
Db 705 TTNKNTS--TETEQANSTISGANVDLOANKOVTFAGSDLUKTTAGNASITGDNVAFVSTEN 762
QY 687 KGLT--TOYRSAGVNF-NGVNGNMSFNLKEGAKVNF---KLKPNENMNTSKPLPIRELA 740
Db 763 KRQTDNTDTTISGGFSYTGVD-----KVGSKADFOYDKQHTQTEVTKNRGSQTEVAG 815
QY 741 NATATGGGVSFFDIYANH--SGRGAELKKEISEINSGANFTLNSHVGGDAPKINKDLTIN 799
Db 816 DLITITANKDLLHEGASHHVEGRYQE---SGENIQHLA--VNDSEYSKTDLSLVNGIDVGN 870
QY 800 AYNSNPSLRQTKDDFYDGY-----ARNAINSTYINISIL-----GGNYTL 838
Db 871 LDYSGVT-KPVKKAIEDGVNTTKPGNNTDLTKKVTAARDAIANLANLSNLETPNVGVEGI 929
QY 839 GQNSSSSITGNITTEKAAAN---VTEANNAPNOQNIIDRVIKLG-SLLVN----- 885
Db 930 KGGGSOQSOTDSQAVSTSNAGKIDISNNKLHDGTHYQSTOEGISLTANTHTSEATLD 989
QY 886 -----GSLSLTGEN--ADIKMLTISESATFKGKTPDTLNTIGNFTNNGT 928
Db 990 KHQTFTHETKGGQGVSTKTSQDITVAIKEGQPTDNALMETKAKGS-----QFTSGND 1044
QY 929 ABINTQGVVVLGNVTNDGDLNITTHAKRNORSIIGDOIINKKGLSNI---TDSNNDAEI 985
Db 1045 ISINVG-----NAHYEG-----AOFDAQK--GKTVINAGDGLTIAQATDTHSESQS 1089
QY 986 QIGGNISOKEGNLTISD-----KINTKQITIKKIDGEDSSSDATSNANLTIKT 1036
Db 1090 NVNGSANLKVGTTPESKDYGGGFNAGTTTHSKEQTTAKVGTITGSQIELNAGHNLTLQG 1149
QY 1037 RELKLITEDLSISGFNKAEITAKDRDLTIGNSNDGNSCAEA---KTVTFNNVKDSKISAD 1093
Db 1150 THLSSEQDIALNATNKKVDLOSASSEHTEKGNLSC--GVQAGFGKKMT-----DDASVN 1202
QY 1094 GHNVTLNS-----KVTSSSNGGRESNDNT--GLTITAKNVENVKDIITSL----- 1138
Db 1203 G-----LGSQAFAIGKQDEKSVREGGTLNNSGNLTINGSVHLQGAQVNSKDTQLTSQS 1258
QY 1139 -----KTVNITASEKVTITTAGSTINATNGKASITT-----KT 1170
Db 1259 DIEITSAQSTDYKNNWGTDIGFNCKKTNTPKEVTEEPATSIHNGKLLVNVVDQOKT 1318
QY 1171 GDISCTISGNTSVSATVDLTITKSGSKTEAKSGEANVT--SATGTIGTGTIS----- 1219
Db 1319 SHQATLETGTLTINSNKDLTL-----SG-ANVTADSVTGNVGSGLSIASQKESDR 1368
QY 1220 -----GNTVNVNTANAGDL-----TVNGNAEINATEGAATLTATGN 1254
Db 1369 HVTGVNVGVNHTDPKSSQVNTAKAGGSLEKTIKDTIDSGIKSSTDAISDKYNSLSS 1428
QY 1255 TLTTEAGSSITSTGQVDLLAQNGSIAGSINAAVNTLTGTLTTVAGSDIKATSGTLVI 1314

```

Db 1429 TIADKTGIS-DETKAKID--QGFGKVGNGIK--NI---VTGAEGHTANADIKVT----- 1474  
QY 1315 NAKDAKLNKGDASDSTEVAVNANSGSVTAATSSVNITGDNTVNGNLNIIISKGRNTV 1374  
Db 1475 -----HVDNDAVTKIT-----SUTSNNDLSLVNGS-TKLITGAEIVSQGQ--V 1515  
QY 1375 RLKCKIEVKYIQPGVASVEEVEIAKRVLEKVKDLSDEBERELAKLGVA--VRFVEPNPT 1433  
Db 1516 DLGGSSVKLENIE-----GHYEGADLDLSSVVD-----LAKOLVGGDISFSPVK 1564  
QY 1434 -ITVNTQ 1439  
Db 1565 NETVNTK 1571  
RESULT 9  
H83135  
Probable adhesin PA4082 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: H83135  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: H83135  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1018 <Sto>  
A:Cross-references: GB:AE004824; GB:AE004091; NID:g9950277; PIDN:AAG07459.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4082

Query Match 7.5%; Score 553.5; DB 2: Length 1018;  
Best local similarity 22.6%; Pred. No. 2.5e-15;  
Matches 301; Conservative 181; Mismatches 453; Indels 399; Gaps 61;  
QY 1 MNKRIYLFKSKRLNALVAVSELARGCDHSTKGE---KPMKVRHLALPLSALLSL 57  
Db 1 MNKYAL-----VNVSGGCVNVVSEGRGRKGPAGAKA-----AIASVALL 43  
QY 58 GVTSI-POSVLASGLOGMDVVGHTATMQV---DGNKTIIRNSVDATII-NMKQFNIDQNM 112  
Db 44 GATAPALAPALPS---GGTVVGGSGANGEIHLGGNSLSVKNQKVDKLIANWDSFVAAGER 100  
QY 113 VQFLOENNSAVENRVTSTNOISOLKGLDSNGOVFLNPNGITIGKDAIINTNGFTASTL 172  
Db 101 VIFNPSSSTALNRVIGTRASDIQGRIDANGOVFLNPNGVLFGRGAQVNVGGLVASTL 160  
QY 173 DIS---NENIKARNTFEOTKDAEIVNH-GLITVKGDSVNLGGKVKHEGVSVN 227  
Db 161 DITDAEFGNSSRYRFTGPTNG-----VLNHGGAITAAEGSIALLGAGVDNRGVLAQ 215  
QY 228 GGSISLLAGOKITSDIINTPIYVIAAPNEAV--NLGDFIAKGGNNVRAAT----- 279  
Db 216 MGVGLGAGSGLTLNFDGNKLLDIRVDAGVANALASNGGLLKADGGRVLMARATANALLN 275  
QY 280 --IRNOGKLSADSVSKDSKNIVL-SAKEGAETGGVISAQ--NQAKGKGLMITGDKVT 334  
Db 276 TVVNSQGAIEARSL-RCKNGRIVLDGPDGKVMVGGALSANALNGPCHGCTVEVRGQAVE 334  
QY 335 LKTCAGVIDLSCKEG--CETVLGGDERGEGKNGIQLAKTSLKSGTINVSKEKGGFAIV 392  
Db 335 VALGTQVNTLASNGLNTWKIAADKID-----VRPSAVSDGVVVAHDTLSRN----- 381  
QY 393 WGDIALIDGNIN-AQSGSDTAKTGGFVETSGHDLFIKDAIVADAKWEHLDFDQVNSINAE 451  
Db 382 -----LASTNIELVSTKGLDLDGVSVAWSGNRL----- 410  
QY 452 PLFNNVTGINDFPTGTGEASDPKKNSELKTTLTNTTISNYLKNWNTITASRKLTVNSS 511

Db 411 -----GLGSAADLTNLGRNAS-----CAKAGLELKAEGA 440  
QY 512 INIGSNHLLHSKG-----QRGGVQIDG--DITSKGNLTIVSGGWVDVHKRNTITDOG 564  
Db 441 IDI--NDKIVLGGAGSALAMDAGEHRVNGTASLAGANATYVSGGY-----YTVQON 493  
QY 565 FLNTAASVAREG-----GNKARDAANAKEYVQAQVTVITTEGKDFRANVSLNCTGKGLN 620  
Db 494 LAQLOAINKNLDGLVYVIGN-----ILGGSYCT-----ALQSIGPAG 532  
QY 621 IISVNNIHLNLSGTINISGNITINQTRKNTSYWQTSHTSHMNVNLSALNLETGANFTFIK 680  
Db 533 VFSG-----TLDGLGNSIGNLSIS-----NTGPNV----- 557  
QY 681 YISSNSKGLTTOYRSSAGVNFNGVNGNMFNLKEGAKVNFKLKPNENMNTSKPLRPLA 740  
Db 558 -----GLFA--RSS-----GTLNKLN-----NLRVSDNTYSGPSSLGALV 593  
QY 741 NITATGGSVFFDIYANHSGRAELKMFSEINISNGANFTLNHVRGDDAFKINKDLTINA 800  
Db 594 GI-----NSGRIANVSVASGVSV--VGSRLRSNALGG---LVGRNISQOI 632  
QY 801 TMSNFSLRQTKDDFYDGYARNAINSTYNISILGGNVTLGGQNSSSSITGNITIEKA-ANV 859  
Db 633 ANASVSGGV-----GVAAS-----TAVGG--LVGENFTTAMGPEAVIENAHNV 675  
QY 860 TLEANNAPNOQNIIRDVTKLGLSLLVNGSLTGENADIKGNLTISESATFKGT---RDT 916  
Db 676 HVAAGSTERN-----LGGVGLVGLNA--KGMIRASGS---QKGVETYPG 717  
QY 917 LNIITG--NETNGTAENITOGVYKLVNVDNDGLNITTHAKRNORSIIIGDIIKKKSL 974  
Db 718 LNVGLGVNMFHSGVSSASQGVAGGAGNTGVLGLSSGGEIPRQASQSVYKSG-- 775  
QY 975 NITDSNDAEIOIGNISQEGNLTISDDKI--NITKQITIKKGDIGDSSDSDATSNANL 1032  
Db 776 -----LATGLIKRAEGNMLGNLAKASGVTDQ-----GGADLGLGVNNSQS 818  
QY 1033 TTKTELKLTEDLSISGNKAEITAKGRDLTIGNSDGNSGAEAKTVTFFNVKDKLSA 1092  
Db 819 ALETAE---ATGKVSQGSNSRVGLIGHNL-----GGSAVAHAISRQDV----- 859  
QY 1093 DGHNTVTLNSKVKTSSNGGSRNSDNDTGLTITAKNVEVNDITSLKTVNITASEKVT 1152  
Db 860 GGFNSLVGLV---GHNGG-----ELNVNDASGRVSAA 889  
QY 1153 AGSTINATNGKASITTKTGDISGTISGNTVSVSATVDLTTSKSKIEAKSGEANYTSATG 1212  
Db 890 ASASV-----GGLVSNAGSILSARSSTVNGSGRIRIGGLVGENQI---QG 933  
QY 1213 TIGGTTISGNTVNVNANAGDLTVNGAEINATEGAATLTATGNTLTTEAGS-SITSTKGV 1271  
Db 934 RIVSSNSECTV---SGDYVYVSMG-----GLAGL-----NLGSIYSVGSCKI 972  
QY 1272 DLLAQN--GSIAGS 1283  
Db 973 DFKPSHYQIYGA 986

## RESULT 10

B99789  
hemagglutinin/hemolysin-related protein [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: B99789  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: B99789  
A:Status: preliminary



Query Match 7.2%; Score 531.5; DB 2; Length 1270;  
Best Local Similarity 22.3%; Pred. No. 2.7e-14;  
Matches 351; Conservative 209; Mismatches 513; Indels 499; Gaps 78;

QY 46 ALKPLSLMALLSLGVTSIPQSIVLSGL--QGMDVVHGTATMQVDGNKTIIRNSVDAIN-W 102  
|| :::: : | : |::| : : | | | | | | | |

```
Db 15 ALTPIALMM-----LSFPVASOAGLVIRKNTVYANGVPVVDINKP-----NGSGLSHNIW 67
QY 103 KQNIIDONENV-----QFOLENN-----SAFNRVTSNOISOLKGLDLSN 143
Db 68 DNLNVKGVVFNNSANESTSLAGNIQSNLTSGSAKVLNEVTSKPNSTINGMEVA 127
QY 144 G---QVFLNPNGITTKDAINTN--GFTASTLDSNENIKARNFTPEOTQDKALAEIV 198
Db 128 GKADLIANPNGITVNGGSGINTKGLTTLTTPGDIOD-----DKLAGYSV 173
QY 199 NHGLITVKGDSVN-----LIGKVKNEGVISVNGGSIISLAGQKITISDIINPTI 249
Db 174 NGGTITLGLDNASPTIELSRNVVAVCKVSADENLVVAGNYYNAAGQ-----V 222
QY 250 TYSIAAPEN-----EAVNLGDIFAKGNNI--NVRAATIRNOGKLS--ADSVSKDKSGNI 299
Db 223 TGSVTSATGRNGYSVDVAKLGGMYANKISLVSTKGVGRNLGVIAGGVNGYSIDSKGNL 282
QY 300 V-----LSAKEGEAEIGGVISAQNOQ-----AKGKMLITGDKVTL 335
Db 283 LNSNAQIOSASTINLTNGTLDNTTGTVTSGTISLNTNKTNTVNTRAGNISTMGD-IVV 341
QY 336 KTGAVIDLSCKEGETVLGGD-----EREGKN-GIOLAKKTSLEKSTINVSKEKG 387
Db 342 NSGTIDTNGKLAAGLAVDTNNTATLNSGKSSGVIE-AGLVALKTGLTNNSQOIRG 400
QY 388 GFALVWGDIADIGNTNAQSGDIATGGFVETSGHDLFTKDNATVDAKEWLLDFDNVSI 447
Db 401 GY-----VGLSNAALN-NNNGDLQITG-----DIALISNGVD-----432
QY 448 NAEDPLENNTGINDPEFTGCE-----ASDPKKNSEKTLTLNTTISNYLKNAMTMITAS 503
Db 433 -----NNKGL---IRKSTGHVIGAAGSVNNGSTRTADTGSSDS-----LGIIA- 473
QY 504 RKLTVNSINIGNSHLILHLSKQORGGVQIDGITSKGNLFIYSGGWVHVHKNITLDO 563
Db 474 -----DTGVIEGAN-----NINNGQ---IASNGVSLSSYSTIIO- 506
QY 564 GFLNITAASVAFEGGNKARDAANAKIVAOGTVTITGEGKDFRANVSLNGTGKGLNIIS 623
Db 507 -----DYAGKILSKSVIKG-----S 523
QY 624 SVNHLTNLSG-----TNISGNITIN-----QTRKNTSVYQWTDHSHVNSALNLETGAN 675
Db 524 SLRNDTGGISGKOGIEVAVGSLTNIGVISSEEGDISLLANSVDNHG-----GFMMGQN 578
QY 676 FTFIKYISSNSKGLTTQYRSAGVNFNGVNGNKSFNLEKAGKYNFKLKPENNMNTSKPLP 735
Db 579 ITM-----ESMSGVNNNTALIVASKKLKINARGSIENROGNFNGAYGL- 622
QY 736 IRLFANITATGGSVFFDIYANHSRGAE-LKMEINISNGANFTLNSHVGRDDAPKINK 794
Db 623 ---YFGMPQQTGGV-----GKEIELSGONIYN-----NSRLIAEDG-----658
QY 795 DLTNATNNSFSLR-----QTKDDPYGYARNAINSTYNTSIILGNNVTLCG---840
Db 659 PLTQAQNTEDNTRALVTSGADASIQVGGTYNNYA-----TWS-----AGNLDIDATTL 709
QY 841 QNSSS-----STGNTTIEKAANTVLEANNAPNOQNRDIRVLIKGLSLLVNGSLSTGEN 894
Db 710 QNSSSGTMDNNAATGFIASDK--NLSLEVYNS-----LNYGWISGKGDVDTVNN 758
QY 895 ADIRKGNLTIS-----ESATFEG--KTRD-----TLNITGNFTNNGTAB-----INI 933
Db 759 GNLYNRNTIAAEKGLDIAALNGIENWKDI SAGGLTMTNTRHVTNNSNSNVGQNVINA 818
QY 934 TOGVVVKLVNTDGLNITTHAK--RNORSIIG-GDIINKKGSNITDSNNDABIQGGNI 991
Db 819 VNDINNRGNIVSDADLVNTTKGNLYNLYVMVGDIALSANSV-----ANNATIEATGDL 874
QY 992 SOKEGNTLISDDKINIKQTIKIGIDGSSDATSNANLTKTKELKLTEDLSISGEN 1051
Db 875 -----IDSKGVNVCNRRGNLHALNGVLSVKGNNLN-NDNGEIRGIG 914
```

---

```
QY 1052 KAEITAKDGRDLTIGNSDNGSGAEAKVTTFNNYKDSKISADGHVNTLNSKVKTSSSNGG 1111
Db 915 -----DVL-----ALGNDYSYKSLTSETG-DVILTANI-----944
QY 1112 RESNSDNTGLTITPAKNVEVN-KDITSLKTVNITASEKVTITAGSTINATNGKASITTK- 1169
Db 945 ---VDNAYGL-IAGENVSVDASTIYNNTALIAANKLVINAGNLENROGNFLNRNG 999
QY 1170 ---TGDISGTISONTVSATVDLTITKSGSKIEAKSGEANVTSATCTIGT-----ISG 1220
Db 1000 ALFGITDNVGGIVGKEGVTLSA--QNVYNNNSIIAENGPLNLS-RETLDNTALLSSG 1056
QY 1221 -----NTVNTANAGDLTVNGAEINATEG-----AATLTATGNTLTTEAGS 1262
Db 1057 ADALIRAAAGTFYNNYATTSYAGNLDVYAASLNNASDGRLEDNTATGVIASOKNLDLSVDN 1116
QY 1263 SIT-----STKQVDDLAAQNGSI-----AGSINAANVTLTNTTGLTITVAGSDIKA 1307
Db 1117 SVTNYGWSIGKGDVHFENVLKGTLVYRNIAADNALTINALNGVENFK---DIVAGTALTI 1173
QY 1308 TSGTLVINAKDAKLNCDASGDSTEVENAVN-ASGSGSVTAATSSSVNITGDLTNTVGNLNI 1366
Db 1174 DFOKVVTN---NSNSNMLGQTIANAVNDINNRGNIVGDIYSLGVKTTG--NIYNLNLML 1227
QY 1367 SKDGRNTVLRGKEIEVKYIQPGVASVEIEAKRVLKVDLSDEERETLAKLGVSAVR 1426
Db 1228 SY-----GVAGV---SANKVTHSGKD-----AVLGGFYGL 1254
QY 1427 FVEPNTTIVNT 1438
Db 1255 ALEANEETNTGT 1266

RESULT 12
T31102
filamentous hemagglutinin 1 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31102
C:Author: J. Bacteriol. 180, 6013-6022, 1998
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326; PMID:9811662
A:Accession: T31102
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-4152 <WAR>
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1
C:Genetics:
A:Gene: lspal

Query Match 7.1%; Score 523.5; DB 2; Length 4152;
Best local similarity 20.1%; Pred. No. 2.6e-13;
Matches 399; Conservative 239; Mismatches 636; Indels 711; Gaps 83;

QY 2 NKIYRLKFSKRLNALVAVSELARGCDHST-----EKSEKPARMKVR-----H 44
Db 3 NKRYKLIFSFKVKNCLVPVAENIKSASGSGSSNSKIAEDQDEEPDLSLACSLSPLSSSIH 62
QY 45 LAL-----KPLSALLSLGVTISIPSVLASGLQGMVDVHGVTATMQVDGNKTII- 92
Db 63 LGLHNSPLKVPFKGKSLSVLLSL---MPATPLLAQQNYAEALNGKVVVDSDSHSSTRIY 118
QY 93 -----RNSVDATII-----NWKOFNIDONEMVQFLOFNNSNAFVNRVTSNQ 132
Db 119 EQKTDNSKDGIVVVEIANPEVDGVSDNRKEFNIP-----NSAVFNSRSTES 166
QY 133 ISOLKGLDLSNGQ-----VELINPNGITIGKDA 160
Db 167 TSOLVGLKHANTLOQKEAKLIILNQVTDGHSNIOGALEAVACKADLIIVNPNGITLNGVK 226
QY 161 IINTNGFTASTLDI--SNENIKARNFTPEOTQDKALAEIVNHGLITVKGDSVNLGKGV 218
```

Db	1173	VLNRTNGTVGVEDKVIYQELNDKAKKEYEDKFAKKFGGRKSRFQNGEFDWAGWAKEGNSY	1173
QY	1116	SONDTGLTITAKNVEVNDITSLKTIVNTASE-KVTTTAGSTIN-	1158
Db	1233	GSKET-----EEKYNGIKKEHTVNIKGKHEIKVPTVSFLENLNNIHQODKSDGIDKSI	1284
QY	1159	-----ATNGKA-----SITT	1168
Db	1285	ISELLAQPIYVAKADVPDVPDPRVAQNDKAVDEGDLGYRTRLSYINQNNYLCAKYFNQLODT	1344
QY	1169	KTGDISG-----TISGNTVSV-----SATV--DIT	1191
Db	1345	EDDKLKGIKRIGDNYFEHOLITRLIERKADVADNHLKLKGLHDIALYKLLDSASIQAKDLN	1404
QY	1192	TKSGS-----KIEAKSGEANTVS-----ATGTI--	1214
Db	1405	LAVGEALTEKQDNLKEDIWYVVTVEVNAQEVLPQVYLAKQTIEEVEKORGVGTGQIRA	1464
QY	1215	-----GGTISGNTVNVTA-----NAGDLTVNGAGAEINTEGAATLTATGNTLT	1257
Db	1465	GIIDVKVDVDRNTGFTIAGYAVGLEAKNKLKNTGDIL---SQRLSKLKVGKGLSESTGTVTV	1521
QY	1258	TEAG-----SSITSTKQGV-----	1271
Db	1522	DETGATKYVKARIKSEGHYYLETDKDNVDLTASELKGNTGQIKAKDLNLDIYETSYKY	1581
QY	1272	---DLLAONG-----SIAGSIN--AANVTLNTT-----	1294
Db	1582	KYEKLFGRNGGEIGDRVTTQSOAKSVGTDSAFDHLHLHLSLGGVNOTGSNLKANRTTGVVK	1641
QY	1295	GLTLTVVAGSDI-----KATSGTLVINAKDAKLNGLDASGDGSTEVNNAVNASGSGSVTAANTS	1348
Db	1642	GDFTNKAKDKLFLHQIDVTTSGT--VYSASASGGQSGAGISLTDQGVETYNKTATAGAN	1699
QY	1349	SSV-----NTGDLNTVNG--LNIISKDRNTVRLRKEKEVEKYIQPGVAS	1392
Db	1700	ADVTNFMKRTRETETSLTHRNSEFNAISLGYLVNGKADIGVDI--ARDVEV-----IKT	1752
QY	1393	VEEVTEAKRVLKVKYDL---SDEGRETLAKLGVSAVRFPENNTITVTNQNETTPRSSQ	1449
Db	1753	PREIAEQKAAEEAKKAEVKEANEASATAAK-----ETEEAENDVAEKDKTKPKFK	1803
QY	1450	VIISE 1454	
Db	1804	KLTD 1808	
RESULT 13			
B81192			
hemagglutinin/hemolysin-related protein NMB0497 [imported] - Neisseria meningitidis			
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001			
C:Accession: B81192			
R:RRTetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.;			
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dough-			
rill, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizz-			
Science 287, 1809-1815, 2000			
A:Authors: Branden, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rapra-			
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B str.			
A:Reference number: A81000; PMID:20175755; PMID:10710307			

A;Residues: 1-1975 &lt;TET&gt;

A; Experimental source: S

1000

## Best Local Similarity

100

Qy 1 MNK-IYRLKSKRNALVAVSELARGCDHSTKSEKPARMKVR-----HLAKPLSA 52  
 Db 1 MNKGHRIIFSKKISTWVAVAEATNSQKKGQAGSSVSLSKTSGLDGLKLTTLKTLVC 60  
 Qy 53 MLLSLGVT-----STP-----QSVLASGLQMDVVHGTATMVDGKNTIIRNSVDA 98  
 Db 61 SLVSLUSVLPAAHAQITTDKSAKPKNOQVVLKNTGAPLVN-----IQTPNGRLSHN----- 112  
 Qy 99 IINMKQFIDONEMVQFLOENNSAVFN---RVTSNQI-----SOLKGILDSNG---VFL 148  
 Db 113 --RYTFQFVDNKGAVLNNDRNNRNNFVVKGSQAQLILNEVRGTFASKLNGIVTVGGQKADVII 170  
 Qy 149 INPNCITIGKDAIINTNGFTASTLIDISNENIKARNFTEQTKKALAEIVNHGIIIT----- 204  
 Db 171 ANPNCIT-----VNGGGF-----KNVGRGILTTGAP 196  
 Qy 205 -VGKDSVNLIGGKVKNEGIVS-----NGG-----SISL-----LAGOKITIS-- 242  
 Db 197 QIGKDA--LTGFDVR--QGLTVGAAGWNGKGGADYTCGLARAVALQCKLOGKNLAVSTG 253  
 Qy 243 ----DINPTITYSTAAPENEAVALNGLDIFAKGN-----INVRATIRNOCKLSADSVSK 293  
 Db 254 POKVDYASGEISAGTRAGTKPTIAL--DPAALGDSITLIANEKGVKNAGTLEA----- 307  
 Qy 294 DKSQNVILSAKGEAEIGVISAQNOQAKGKLMITGDKVTLKYGAVIDLKSGEGETYL 353  
 Db 308 -----AKQ-----LIVTSSGRIENSGRIATTAD-----GTEASPTYL 339  
 Qy 354 GDERGEGKNGIOLAKTSLKSGSTINVSKEKGFALVWG--DIALIDGNIAQSGDIA 412  
 Db 340 STETKEKAAG-----TFISNGGRI-----ESKGLLVITETGEDISLRNGAV--VQNGG--S 386  
 Qy 413 KTGGEVETSGHDLFIKDNAIDVAKEWLLDFONVSINAEDPLF-----NNTGIND--EFTGT 467  
 Db 387 REATTVLNAGHLVIESKTNVN-----NAKGPATLSADGRTVIKEASIQRT 433  
 Qy 468 GEASDPKKNSEL-----KTLTNTTISN--YLNKAMTMTITASRKL-----TVNS 510  
 Db 434 TVYSSSKGNAELGNTRITGADTVLSNGTISSSAVIDAKDTAIEAGKPLSEASTVTS 493  
 Qy 511 SNIGNSHLILHSGKGGGVQIDGDTISK-----GNLTIYSGWVDVHKNTLTDQ 564  
 Db 494 DRLNGGS-----IKGKQALADDNITAKTNTLNPGLYVHTG--KDLNLDVKD-- 544  
 Qy 565 FLNITAAVAFEGGNKARDANAKIVAQGVITIGCKDFRANVSLNGTKGKGLNISS 624  
 Db 545 ----LSAASHLKSDN-----AAH-----ITGTSKLTASK--DMGVAGSLNVTNT 585  
 Qy 625 VNNLTHNLSGTINI---SGNITINOTRKNTSYWTSHDSHWNVSALNLETGANFTFYK 681  
 Db 586 --NLRTN--SGNLHIQAAKGNLQI--RNTKLN-----AAKALETTA----- 620  
 Qy 682 ISSN--SKGLTQYRSSAGVNFNG-----VNGNMSFNLKEGAKVNFKLKPNENMTSKPL 734  
 Db 621 LQGNIVSDGL-----HVSADGHRVSLIANGNADET-----GHNTLTAKA-- 659  
 Qy 735 PIRFLANITATGGGSVEFFDIYAHNSGGAELKMWSEINSGANFTLNSHVRGDDAFKINK 794  
 Db 660 -----DYNAGSVGK--RLKADNTNITSSG----- 683  
 Qy 795 DLTINATNSFSLRQTKDFDGYARNAINSTY--NLSILGGVNTLGGON--SSSITGNIT 852  
 Db 684 DITLVAGN-----IQLDGQKQNSLNGKHISIKNNGNADLKNLVHAKSALNIH 735  
 Qy 853 IEKA---ANVTLEA---NNAPNOQNI-----DRVILKGLSLL 883  
 Db 736 SDRALSIENTKLESTHNTHLNAQHERHTLQVDAVHRHLSITGSOIWNOKLPSANKLV 795  
 Qy 884 VNGSLTGENADIKGNLTISESATPKGTRDTLNTGNFTNNGTAENITQG---VVKLG 941  
 Db 796 ANGVALNARYSQIADNTTLRAGA-----INLTAGTALVVRG 832  
 Qy 942 NVTNDGDLNITTHAKRNQRNSIIGGDIINKKGSINIT-----DSNNDAEIQIGGN--IS 992

Db 833 NI--NWSTVSTKLTLEADNAELKPLAGRLNIEAGSGTLTIEPANRISATHTDLSIKTGKLLLS 891  
 Qy 993 OKEGNLTISDDKINIKQITIKKIDGEDS--SSDATSNANLTIKTKELKLTEDLSISGF 1050  
 Db 892 AKGNAGAPSAQVSSLEAKGNIRLVGTGEDLRGSKITAKGNLVVATTKKLNEAVNNSF 951  
 Qy 1051 NK-----AETAKOGRDLTICNSDNGSNGAEPKVTNNVVDKSLADGHNVTLSKV 1103  
 Db 952 SNYFPTQKAALKNOK--SKELEQOIAOLKKSSPKSLIP--TLOEERDLAFYIOAINKEV 1008  
 Qy 1104 KTSSSNG-----GRESNDNDTGLTITAKNVEYN--KDIITSLKTVNITASEKVTITTAGSTI 1157  
 Db 1009 GKPKPKGKEYLOAKLSAQNID--LISAQGIETISGSDITASKKLNHLAAAGVLPKAADSEA 1065  
 Qy 1158 -----NATNGKASITTKTG-----DISGTISG----- 1179  
 Db 1066 AAILDGIITDQYEIGKPTYKSHYDKAALNKPRLTGTGTGVSIIHAAAALDDARIIGASEI 1125  
 Qy 1180 -----NTVSVSATVDLTTKSGS-----KTEAKSG-----EANV 1207  
 Db 1126 KAPSGSIDIKAHSDIIVLEAGONDAYTLTKGSGKIIRKTKFTSTRDHLIMPAPVELTA 1185  
 Qy 1208 TSATGTIGTISGNTVNTANAGDLTVNGCAEIN--ATEG-----AATLTATGNLTTEAG 1261  
 Db 1186 NGITLQAGNIEANTTFNAPAGKVTYLVAGEELQALLAEEGIHKHELDVQKSRFFIGKVG 1245  
 Qy 1262 SS-----ITSTKGOVDLLAONGSIAGSIANAANVTNTTCTLTITVAGSDIKATSGTLVIN 1315  
 Db 1246 KSNYSKNEMLNETKLPVRVAAQT---AATRSGWDTVLTEGTEFTKLTAGADLOAGVEKA-- 1300  
 Qy 1316 AKDAKL-----NGDASGDSSTEVNAV---NAGSGSV--TAATSSSVNITGDLNVTNGLN 1364  
 Db 1301 RADAKIILKLGIVNRIOSEEKLETNSTVWQOAGRGSTIETLKLPSFSPPTKLTPAGGY 1360  
 Qy 1365 I--ISKDGRNTVRLRGKEIEVKYIQPGVASVEEVEIAKRV---LEKVKDLSDEERETLA 1418  
 Db 1361 IVDIPKGNLKT-----EIEKLAQOPEYAYLKQLOVAKNWNQVQVLAQDKWDYKQEGLT 1414  
 Qy 1419 KLGVSAVRVF 1428  
 Db 1415 RAGAAIVTII 1424

RESULT 14  
 AC0304  
 Probable hemolysin YPO2490 [imported] - Yersinia pestis (strain C092)  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 R:Accession: AC0304  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barril  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AC0304  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2535 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: YPO2490

Query Match 7.08; Score 520; DB 2; Length 2535;  
 Best Local Similarity 19.7%; Fred. No. 1.9e-13;  
 Matches 389; Conservative 272; Mismatches 636; Indels 642; Gaps 85;  
 Qy 1 MNK-IYRLKSKRNALVAVSELARGCDHSTKSEKPARMKVRHLAKPLS-AMLSLG 58  
 Db 5 MNKLYRIIFNKVRGMIVVADIAASGRSSSPSGLGHTQHRIRISALTSLSLLALG 64  
 Qy 59 VTS-----IPQSVLASGLQGMVYV---HGT-----ATMQVDGNKTI 91

Db	1096	LNNWRGSDYVLOQLNNDP	VFKR-----LCDNAYEQRLVRDQVLAKTQGVASDY--RSA	1149
QY	811	KODFYDGYARN-AIN	STYNISILGGNVTLGGQSSSITGNITIEKAANVTLCANNAPNQ	869
Db	1150	OEQFEALFAAGLEYSKAFNIA-LGTHLS---	EQMAALTHNI-----VLMETRDVAGO	1198
QY	870	QNIRDRV-----IKGLSLLVNGSLSGENADI---	-----KGNLTISSEATPKGTRDTLL	917
Db	1199	TVLVPVYLAGVKPGDLQALANG-IAEAENISL	TEVQFTNAGAITATNDLKISMAQDITL	1257
QY	918	NITGFTNNGTAEINITQGVKVLGNVTDGDLNITHAKRNQBSI---	-----TGGDIINKKGS	974
Db	1258	NNRGGLQAG-----GDMQLSTLNSDIDL---	TSARINATNLQDSGRDVLRTDSA	1306
QY	975	NITDSNDAIEIGGNISQKE---GNLTJSSKINIKOITTKK	IGID-----GB	1021
Db	1307	QLSSDN-----GAVSRDQITILGLA---	SINVNNAITNGRDFIMQGASLNVGDO	1354
QY	1022	SSSDATSNANL-TITKTELKLTEDUSISGFNKAETAKD-----	GRDLTIGNSDNGNSGAE	1076
Db	1355	LQVTTGGDWLQETVQTRDQISHTD-----GRGSATSEH	IRLHGSEVNVNGGALTAN--VD	1406
QY	1077	AKTVTFNVNVDKSIKISADGHNVITLNSKVK-----	TSSNGGRESNDNDGLTITAKNVEYN	1132
Db	1407	NUTVAGANINAAITLVEQAQNISLSAATDSLHVITGESSKRRHTSSV	NYDETLGSLQLNAT	1466
QY	1133	KDITSLKTVNITASEKVTTTAGSTTNATNGKASITTKT-----	-----GDISGHTS	1178
Db	1467	GDINILOAAQDITILRASAVQITDGAALTAAGGDVLLTTQTQEH	DEQRNHTGLSKGIASSTLT	1526
QY	1179	GNTVSATVDLTTKSGKIEAKSGEANTVSAITGIGT--	ISGNTVNVTANAGDLTVGN	1236
Db	1527	RTEDLSQTLAV---GSMLSA--GSDVSGKNIAVMGSNVVADQISL	RAQE-NITVGT	1579
QY	1237	GAE-----INAYEGAATLTATGNLTLTTEAGSSTTPKQGVODLLA	QNGSTAG	1282
Db	1580	AQOSESESHLFEQKSGLMSTGGICVTVCSSTKMTDSQSISSV-----	-----GSTVG	1629
QY	1283	SINAAENVLTNTGTLTTTVAGSDIKATSTLYINAKDAKLNGD-----	ASGDSTEVNAVN	1336
Db	1630	SV-----LGNVSWTAGEDLRL-VQAEVLAKDINLTGNKNVSILAENOL	TOSHTE	1679
QY	1337	ASGSG-----SVTAATSSVNITGDNLTVNGLINLISKDGRNVT	RLRGKEI	1381
Db	1680	QKQSGLLHALSGAVGSANVTAVTTAKAASESSGRKLQALQ-----	-----	1720
QY	1382	EVKIYQPCVASVE--EVTEAK-----RVLEKVKOLSDSEERETLAK	LGVSAVRFV	1428
Db	1721	-VKAALNGVQVQAGVLQVABEGGDAAMFGISASLGSKQSSSEQHQETHV	TGST----	1775
QY	1429	EPNNTIIVNTQNEFTTRPSSQVVIIESEKACFSSGN	1463	
Db	1776	TAGNNLTINATGEGNAANSGDIVY-QGSQLOAQGD	1809	

hemagglutinin/hemolysin-related protein NM81779 [imported] - *Neisseria meningitidis* C:Species: *Neisseria meningitidis*  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: G81044  
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizaa, M. Science 287, 1809-1815, 2000  
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. A:Reference number: AB1000; UID:20175755; PMID:10710307  
A:Accession: G81044  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1995 <TEF>  
A:Cross-references: 1-1995 <TEF>





